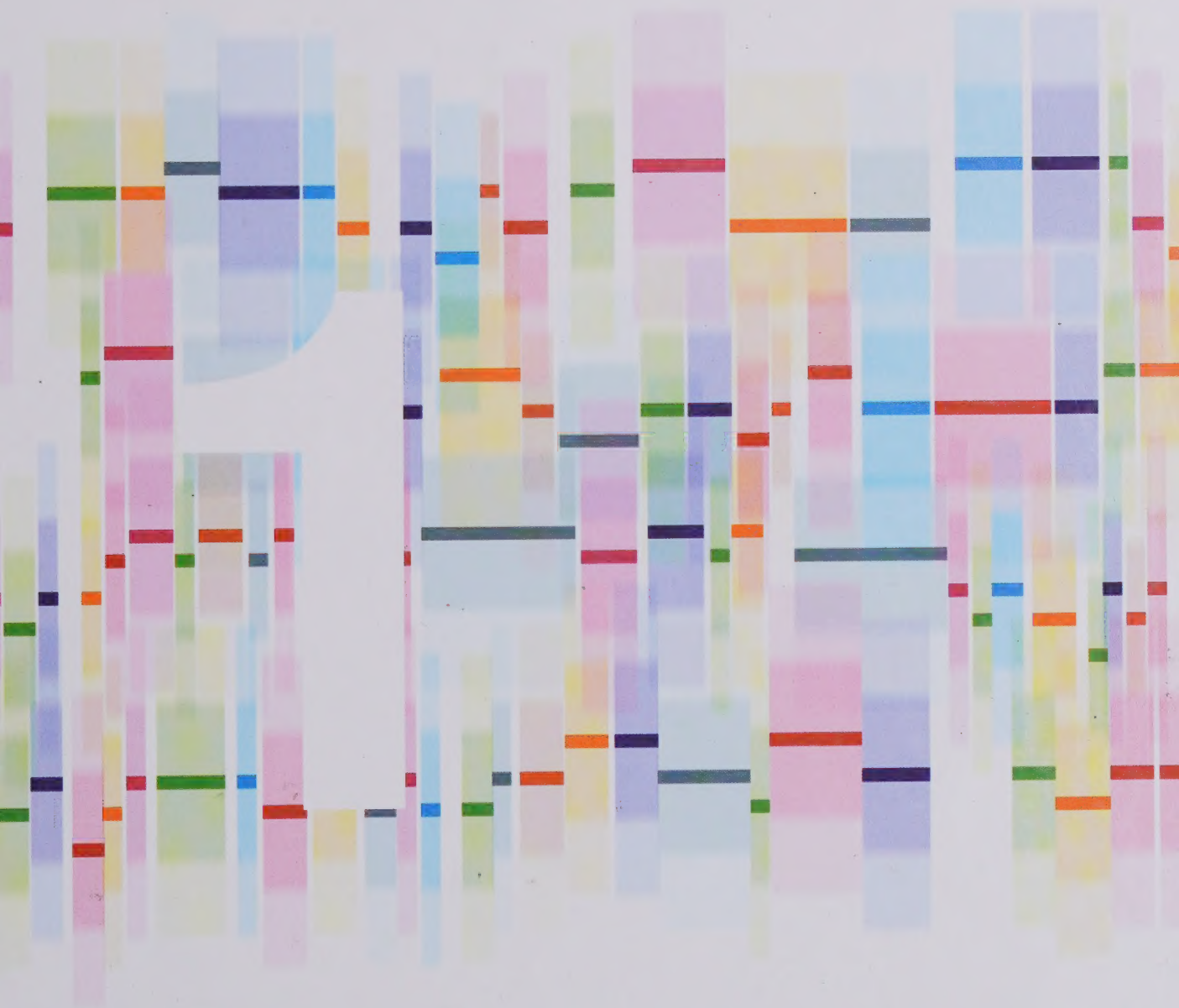


Volume 1

ICRF Handbook of Genome Analysis

Nigel K. Spurr
Bryan D. Young
Stephen P. Bryant



Tara C. Matise
Rockefeller University
1230 York Ave, Box 192
New York, NY 10021

ICRF Handbook of Genome Analysis

Genome Analysis

Section 1

Michael A. Hahn

University of California, San Diego

Section 2

Section 3

Section 4

Section 5

Section 6

Section 7

Section 8

Section 9

Section 10

Section 11

Section 12

Section 13

Section 14

Section 15

Section 16

Section 17

Section 18

Section 19

Section 20

Section 21

Section 22

Section 23

Section 24

Section 25

Section 26

Section 27

Section 28

Section 29

Section 30

Section 31

Section 32

Section 33

Section 34

Section 35

Section 36

Section 37

Section 38

Section 39

Section 40

Section 41

Section 42

Section 43

Section 44

Section 45

Section 46

Section 47

Section 48

Section 49

Section 50

Section 51

Section 52

Section 53

Section 54

Section 55

Section 56

Section 57

Section 58

Section 59

Section 60

Section 61

Section 62

Section 63

Section 64

Section 65

Section 66

Section 67

Section 68

Section 69

Section 70

Section 71

Section 72

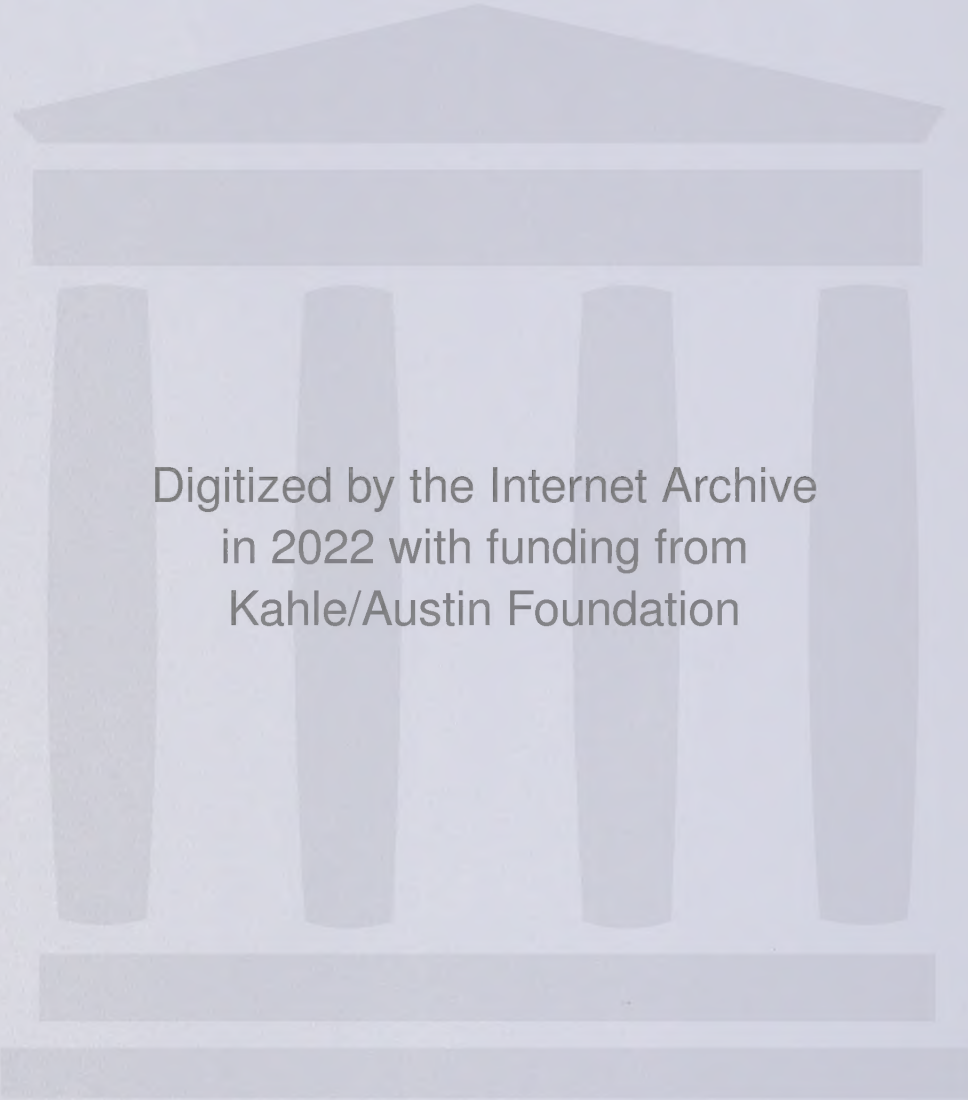
Section 73

Section 74

Section 75

Section 76

b



Digitized by the Internet Archive
in 2022 with funding from
Kahle/Austin Foundation

ICRF Handbook of Genome Analysis

THE EDITORS

Nigel K. Spurr

*SmithKline Beecham Pharmaceuticals
New Frontiers Science Park
Harlow, Essex*

Bryan D. Young

*ICRF Department of Molecular Oncology
St Bartholomew's Hospital
London*

Stephen P. Bryant

*Gemini Research Ltd
Science Park
Cambridge*

THE SECTION EDITORS

Stephan Beck

*The Sanger Centre
Hinxton, Cambridge*

Anna-Maria Frischauf

*ICRF
Lincoln's Inn Fields
London*

Denise Sheer

*Human Cytogenetics Laboratory
ICRF, Lincoln's Inn Fields
London*

In two volumes

Volume 1

b

**Blackwell
Science**

The content of this book expresses the views of the individual authors and not those of the Imperial Cancer Research Fund as a registered UK charity

© 1998 by
Blackwell Science Ltd
Editorial Offices:
Osney Mead, Oxford OX2 0EL
25 John Street, London WC1N 2BL
23 Ainslie Place, Edinburgh EH3 6AJ
350 Main Street, Malden
MA 02148 5018, USA
54 University Street, Carlton
Victoria 3053, Australia

Other Editorial Offices:
Blackwell Wissenschafts-Verlag GmbH
Kurfürstendamm 57
10707 Berlin, Germany

Blackwell Science KK
MG Kodenmacho Building
7-10 Kodenmacho Nihombashi
Chuo-ku, Tokyo 104, Japan

All rights reserved. No part of
this publication may be reproduced,
stored in a retrieval system, or
transmitted, in any form or by any
means, electronic, mechanical,
photocopying, recording or otherwise,
except as permitted by the UK
Copyright, Designs and Patents Act
1988, without the prior permission
of the copyright owner.

First published 1998

A catalogue record for this title
is available from the British Library

ISBN 0-632-03728-8

Library of Congress
Cataloging-in-publication Data

ICRF handbook of genome analysis /
edited by Nigel Spurr.

p. cm.

Includes bibliographical references
and index.

ISBN 0-632-03728-8

1. Gene mapping-Laboratory
manuals. 2. Nucleotide sequence-

Laboratory manuals. I. Spurr,
Nigel.

QH445.2. I27 1996

574.87' 3282-dc20

96-17981

CIP

Set by Setrite Typesetters, Hong Kong
Printed and bound in Great Britain by
MPG Books Ltd, Bodmin, Cornwall

The Blackwell Science logo is a
trade mark of Blackwell Science Ltd,
registered at the United Kingdom
Trade Marks Registry

DISTRIBUTORS

Marston Book Services Ltd
PO Box 269
Abingdon
Oxon OX14 4YN
(Orders: Tel: 01235 465500
Fax: 01235 465555)

USA

Blackwell Science, Inc.
Commerce Place
350 Main Street
Malden, MA 02148 5018
(Orders: Tel: 800 759 6102
781 388 8250
Fax: 781 388 8255)

Canada

Copp Clark Professional
200 Adelaide St West, 3rd Floor
Toronto, Ontario M5H 1W7
(Orders: Tel: 416 597-1616
800 815-9417
Fax: 416 597-1617)

Australia

Blackwell Science Pty Ltd
54 University Street
Carlton, Victoria 3053
(Orders: Tel: 3 9347 0300
Fax: 3 9347 5001)

.....

Contents

.....

Volume 1

List of protocols, IX

List of contributors, XIII

Preface, XVII

Sir Walter Bodmer

Introduction, XIX

Nigel K. Spurr

Section 1: Genetic mapping

Section editor: *Nigel K. Spurr*

Introduction, 3

Nigel K. Spurr

- 1 Genetic mapping: family studies, 5**
Sarah P. West
- 2 Genetic analysis of complex traits, 27**
D. Timothy Bishop
- 3 Constructing and using genetic maps, 43**
Stephen P. Bryant
- 4 Automated map construction: MultiMap, 89**
Tara Cox Matise & Aravinda Chakravarti
- 5 Identification and characterization of DNA polymorphisms, 97**
John A.L. Armour

6 DNA typing, 127

John A.L. Armour

Section 2: Physical mapping 1: cytogenetic analysis

Section editors: *Bryan D. Young & Denise Sheer*

Introduction, 147

Bryan D. Young

- 7 Conventional cytogenetics, 149**
Barbara Czepulkowski
- 8 Solid tumour cytogenetics, 187**
Sandra Birdsall, Yong-Jie Lu & Janet Shipley
- 9 Fluorescence in situ hybridization, 213**
Gabriele Senger & Denise Sheer
- 10 Chromosome painting, 241**
Lyndal Kearney
- 11 Chromosome microdissection, 263**
Debra M. Lillington & Andrew N. Shelling
- 12 Chromosome sorting and analysis by fluorescence-activated flow sorting, 291**
Simon P. Monard
- 13 Digital microscopy for fluorescence in situ hybridization, 305**
Nigel P. Carter

Section 3: Physical mapping 2: long-range mapping and gene isolation

Section editors: *Anna-Maria Frischauf & Nigel K. Spurr*

Introduction, 319

Anna-Maria Frischauf

- 14 Somatic cell hybrid approaches to genome mapping, 321

Alan J. Schafer & Christine J. Farr

- 15 Long-range physical mapping, 367

Sebastian Meier-Ewert, Leonard Schalkwyk, Fiona Francis & Hans Lehrach

- 16 Long-range physical map construction and the integration of genetic and physical maps, 421

Richard Mott, Andrei Grigoriev & Hans Lehrach

- 17 Strategies for rapid isolation of genes from genomic DNA: cDNA techniques, 441

Michael A. North, Bernhard Korn, Roger Cox & Hans Lehrach

- 18 Screening cDNA libraries by transient expression in mammalian cells, 469

David L. Simmons

- 19 Denaturing gradient gel electrophoresis and its variants in mutation detection, 495

Rob B. van der Lijst & Riccardo Fodde

Index, I

Colour plates fall between pp. 262 and 263

Volume 2

List of protocols, IX

List of contributors, XIII

Preface, XVII

Sir Walter Bodmer

Introduction, XIX

Nigel K. Spurr

Section 4: DNA sequencing

Section editor: *Stephan Beck*

Introduction, 513

Stephan Beck

- 20 Sequencing strategies, 517

Carol Churcher, Mary Berks, Sharen Bowman, David Buck & Karen Thomas

- 21 Template amplification and purification, 529

André Rosenthal, Matthias Platzer, Gerald Nyakatura & Monique Jouet

- 22 Sequencing chemistries: chemical vs. enzymatic DNA sequencing and conventional and novel technologies, 557

Peter Heinrich & Horst Domdey

- 23 Electrophoresis, 577

Christoph Heller

- 24 Sequence labelling and detection, 599

Peter Richterich

- 25 Repeat analysis, 617

Aleksandar Milosavljević

Section 5: Other model systems

Section editor: *Nigel K. Spurr*

Introduction, 631

Nigel K. Spurr

- 26 Mouse genome mapping, 633

G. Argyropoulos & S.D.M. Brown

- 27 Gene therapy, 649

Richard G. Vile & Rosa Maria Diaz

- 28 Drosophila genome maps, 667

Robert D.C. Saunders

- 29 The genome of *Caenorhabditis elegans*, 687

John Sulston, Robert H. Waterston and members of the Caenorhabditis elegans Genome Consortium

- 30 The yeast genome project, 695

Horst Feldmann

- 31 Bacterial genomes: the *Escherichia coli* model, 715

Millicent Masters & Guy Plunkett III

- 32 Problems in plant genome analysis, 745

Prem P. Jauhar

- 33 The Arabidopsis nuclear genome, 761

Michel Delseny & Richard Cooke

- 34 Rice as a model for genome analysis, 789

Akio Miyao, Kimiko Yamamoto, Ilkka Havukkala, Masahiro Yano, Nori Kurata, Yuzo Minobe & Takuji Sasaki

Section 6: Internet resources

Section editor: *Stephen P. Bryant*

Introduction, 819

Stephen P. Bryant

35 Internet resources and how to use them, 821
Martin J. Bishop

36 UNIX system survivors' guide, 837
Jaime Prilusky

37 Databases and genome resource centres, 843

Appendices

I Materials, media and solutions, 863

II Preparation of blood bottles and processing of
blood samples, 869

III Commercial suppliers, 873

IV Fluorochromes and filter sets for FISH, 877

V Useful addresses and Internet connections, 879

VI Basic data on human and other genomes, 887

VII Catalogue of mapped human disease genes,
893

VIII Mouse gene knock-out tables, 943

IX Chromosome aberrations associated with
cancer, 977

Abbreviations, acronyms and glossary, 985

Index, I

List of protocols

Volume 1

- 1 Using IGD and X-PED to assist in a linkage analysis, 54
2 Creating input files for LINKAGE, 57
3 Modified procedure for Protocol 2 if a lod score table is required, 60
4 An illustrative session with CRI-MAP, 64
5 Pinpointing a disease gene with LINKMAP, 83
6 Pre-association of dispersed repeat elements, 114
7 SSCP analysis for polymorphism, 115
8 Typing amplified RFLPs (PCR-RFLPs), 118
9 Typing dinucleotide repeat loci, 119
10 DNA fingerprinting, 136
11 Typing using locus-specific minisatellite probes, 138
12 MVR-PCR at D1S8, 140
13 Processing of whole blood samples and slide preparation, 168
14 Thymidine block synchronization method for obtaining prometaphase chromosomes, 171
15 Fragile X detection, 172
16 Setting up amniotic fluid cultures, 173
17 Setting up a chorionic villus sample, 174
18 Setting-up a long-term villus culture, 175
19 Direct villus culture, 176
20 Harvesting amniotic fluid and chorionic villus cultures, 177
21 Giemsa banding, 179
22 Quinacrine banding, 180
23 Reverse banding, 182
24 Constitutive heterochromatin banding, 183
25 Synchronization technique for bone marrow samples, 184
26 Mechanical disaggregation, 196
27 Enzymatic disaggregation, 197
28 Direct chromosome preparations, 198
29 Establishing short-term cultures from cell suspensions, 199
30 Establishing short-term cultures from explants, 201
31 Lipofectin-mediated transfection for establishing long-term cultures, 202
32 Passaging cells, 203
33 Harvesting by removal of adherent cells, 204
34 Harvesting in situ, 205
35 Labelling genomic DNA by nick translation for CGH and hybridization for CGH, 206
36 Tumour imprint preparation, 209

- 37 Pretreatment of tumour imprints prior to hybridization, 209
- 38 Pre-banding with Wright's stain, 222
- 39 BrdU incorporation during late S-phase for replication G-banding, 223
- 40 BrdU incorporation during early S-phase for replication R-banding, 224
- 41 The release of chromatin from interphase nuclei, 225
- 42 DNA isolation from agarose plugs, 226
- 43 Alu-PCR with primer AGK34, 227
- 44 DNA labelling with biotin and digoxigenin by nick translation using a commercial kit, 228
- 45 Quality control of biotin and digoxigenin labelling, 229
- 46 Denaturation of chromosomal DNA, 231
- 47 Denaturation and prehybridization of probe DNA and hybridization to chromosomal DNA, 232
- 48 Posthybridization washes, 233
- 49 Detection of hybridized probes, 234
- 50 Nick translation, 250
- 51 Alu-PCR amplification of flow-sorted chromosomes, 252
- 52 DOP-PCR amplification of flow-sorted normal and abnormal chromosomes, 253
- 53 Competitive (or chromosomal) in situ suppression hybridization, 255
- 54 Detection of hybridized labelled probe, 257
- 55 Lymphocyte separation of peripheral blood samples, 274
- 56 Cryopreservation of blood and bone marrow samples, 275
- 57 Unsynchronized cultures, 276
- 58 Harvesting chromosomes for microdissection and microcloning, 277
- 59 Preparation of microneedles, 278
- 60 Chromosome microdissection, 279
- 61 DOP-PCR amplification reaction, 280
- 62 Fluorescence in situ hybridization using microdissected region-specific probes, 282
- 63 Microcloning of DOP-PCR amplification products, 286
- 64 Colony PCR to isolate microclone inserts, 287
- 65 Preparation of chromosomes by the polyamine method for flow sorting, 299
- 66 Chromosome preparation by the magnesium sulphate method, 301
- 67 Preparation of chromosomes for library construction, 302
- 68 Preparation of HAT medium, 346
- 69 Whole-cell fusion of mammalian cells, 347
- 70 Optimization of micronucleation, 350
- 71 Enucleation from plastic bullets, 351
- 72 Percoll gradient enucleation: an alternative protocol for enucleation, 354
- 73 Filtration of microcells, 354
- 74 Fusion of microcells to whole recipient cells, 355
- 75 Production of radiation hybrids, 358
- 76 Preparation of high molecular weight DNA in agarose from human cell lines or mouse spleens, 378
- 77 High molecular weight liquid DNA preparation, 380
- 78 YAC library construction, 381
- 79 Arraying YAC libraries in microtitre plates, 389
- 80 YAC library filter lifts, 389
- 81 Preparation of YAC agarose blocks, 391
- 82 Size fractionation of YACs by pulsed-field gel electrophoresis, 393
- 83 Partial restriction digest mapping of YAC DNA, 394
- 84 Generation of end-specific probes from YAC clones, 395
- 85 Cosmid library construction, 397
- 86 P1 library construction, 401
- 87 Extraction of DNA from P1 clones: minipreps, 414
- 88 Extraction of DNA from P1 clones: maxipreps, 415
- 89 Competition of probes to remove repetitive sequences, 416
- 90 Hybridization and washing, 417
- 91 Exon trapping by pSPL1 from genomic DNA cloned in cosmids, 449
- 92 Direct cosmid or YAC hybridization to cDNA filters, 456
- 93 cDNA enrichment using solution hybridization to genomic clones as probes, 458
- 94 Rapid amplification of 3' ends, 462
- 95 5' RACE using a unique oligonucleotide ligated to the 3' end of the first-strand cDNA, 464
- 96 cDNA library construction, 478
- 97 Screening for cell-surface proteins by transient expression, panning and episomal rescue, 486
- 98 Functional adhesion assays on cloned cDNAs transiently expressed in COS cells, 491

- 99 Denaturing gradient gel electrophoresis, 503
- Volume 2**
- 100 Standard alkaline lysis miniprep of plasmid DNA, 539
- 101 PEG precipitation of plasmid DNA, 540
- 102 Short alkaline miniprep for DNA, 540
- 103 Standard PEG–phenol method for recovery of DNA from M13 phage, 541
- 104 Detergent extraction method for M13 DNA (for 96 samples), 542
- 105 Magnetic bead purification of M13 DNA (for 96 samples) and bead re-use, 543
- 106 Silica gel membrane (or glass filter) purification of M13 DNA (24 samples), 545
- 107 Preparation of phagemid DNA, 546
- 108 CsCl gradient purification of cosmid DNA, 546
- 109 Generation of single-stranded DNA sequencing template by asymmetric PCR (one-step), 548
- 110 Generation of single-stranded DNA sequencing template by asymmetric PCR (two-step), 549
- 111 Generation of double-stranded DNA sequencing template by symmetric PCR (for 24, 48 or 96 templates), 549
- 112 Purification of PCR-generated double-stranded DNA sequencing template by selective polyethylene glycol (PEG) precipitation, 551
- 113 Generation of single-stranded DNA sequencing template by PCR followed by affinity capture using the biotin–streptavidin system, 551
- 114 Recovery of PCR product by the ‘freeze and squeeze’ method, 552
- 115 Recovery of PCR product by column purification, 553
- 116 Recovery of PCR product by the agarose method, 554
- 117 Direct sequencing of DNA in low-melting-point agarose, 555
- 118 How to set up and run a standard sequencing gel, 588
- 119 How to set up and run a direct blotting gel, 590
- 120 How to set up and run a sequencing gel for automated sequencing, 592
- 121 How to set up and run a capillary electrophoresis gel for sequencing, 593
- 122 Labelling oligonucleotides with NHS- or ITC-haptens, 612
- 123 Preparation and purification of oligonucleotide–alkaline phosphatase conjugates, 613



List of contributors



G. Argyropoulos *Department of Biochemistry and Molecular Genetics, St Mary's Hospital Medical School, Imperial College of Science, Technology and Medicine, London W2 1PG, UK*

John A.L. Armour *Department of Genetics, School of Medicine, Queen's Medical Centre, Nottingham NG7 2UH, UK*

Mary Berks (deceased) *The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*

Sandra Birdsall *Molecular Cytogenetics Laboratory, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK*

D. Timothy Bishop *ICRF Genetic Epidemiology Laboratory, St. James's University Hospital, Leeds LS9 7TF, UK*

Martin J. Bishop *HGMP Resource Centre, Hinxton, Cambridge CB10 1SB, UK*

Sharen Bowman *The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*

S.D.M. Brown *MRC Mouse Genome Centre, Harwell, Oxfordshire OX11 0RD, UK*

Stephen P. Bryant *Gemini Research Ltd, 162 Science Park, Milton Road, Cambridge CB4 4GH, UK*

David Buck *The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*

Nigel P. Carter *The Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1RQ, UK*

Aravinda Chakravarti *Department of Genetics, Room BRB 721, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA*

Carol Churcher *The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*

Richard Cooke *Laboratoire de Physiologie et Biologie Moléculaire des Plantes, URA 565 du CNRS, Université de Perpignan, 66860 Perpignan, France*

Roger Cox *ICRF, Wellcome Trust Centre for Human Genetics, Nuffield Department of Clinical Medicine, Windmill Road, Headington, Oxford OX3 7BN, UK*

Barbara Czepulkowski *Department of Cytogenetics/Haematology, Kings College Hospital, Denmark Hill, London SE5 9RS, UK*

Michel Delseny *Laboratoire de Physiologie et Biologie Moléculaire des Plantes, URA 565 du CNRS, Université de Perpignan, 66860 Perpignan, France*

Rosa Maria Diaz *Richard Dimpleby Department of Cancer Research, Rayne Institute, St Thomas' Hospital, Lambeth Palace Road, London SE1 7EH, UK*

Horst Domdey *Laboratoire für Molekulare Biologie, GenZentrum, Ludwig-Maximilian-Universität, Am Klopferspitz, 8033 Martinsried, Germany*

Christine J. Farr *Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK*

Horst Feldmann *Adolf-Butenandt-Institut für Physiologische Chemie, Physikalische Biochemie und Zellbiologie der Universität, Schillerstrasse 44, D-80336 München, Germany*

Riccardo Fodde *MGC-Department of Human Genetics, Sylvius Laboratory, Faculty of Medicine, University of Leiden, Leiden, The Netherlands*

Fiona Francis *Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin, Germany*

Andrei Grigoriev *Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin, Germany*

Ilkka Havukkala *Rice Genome Research Program, Institute of Society for Techno-Innovation of Agriculture, Forestry and Fisheries, 446-1, Ippaizuka Kamiyokoba, Tsukuba, Ibaraki 305, Japan*

Ulrich Hentschel *abGene GmbH, Lochhamer Strasse 11, D-14195 Berlin, Germany*

Christoph Hees *Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin, Germany*

Prem P. Jauhar *USDA-ARS, Northern Crop Science Laboratory, Fargo, North Dakota 58105-5677, USA*

Monique Jouet *University of Cambridge Clinical School, Addenbrooke's Hospital, Department of Medicine, Hills Road, Cambridge CB2 2QQ, UK*

Lyndal Kearney *MRC Molecular Haematology Department, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK*

Bernhard Korn *ICRF, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK*

Nori Kurata *Rice Genome Research Program, National Institute of Agrobiological Resources, 1-2 Kannondai 2-chome, Tsukuba, Ibaraki 305, Japan*

Hans Lehrach *Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin, Germany*

Debra M. Lillington *ICRF Department of Medical Oncology, Science Building, St Bartholomew's Hospital Medical College, Charterhouse Square, London EC1M 6BQ, UK*

Yong-Jie Lu *Molecular Cytogenetics Laboratory, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK*

Millicent Masters *Institute of Cell and Molecular Biology, The University of Edinburgh, Darwin Building, King's Buildings, Mayfield Road, Edinburgh EH9 3JR, UK*

Tara Cox Matise *Department of Statistical Genetics, The Rockefeller University, 1230 York Avenue, Box 192, New York, NY 10021, USA*

Sebastian Meier-Ewert *Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin, Germany*

Aleksandar Milosavljević *CuraGen Corporation, 555 Long Wharf Drive, New Haven, CT 06511, USA*

Yuzo Minobe *Rice Genome Research Program, National Institute of Agrobiological Resources, 1-2 Kannondai 2-chome, Tsukuba, Ibaraki 305, Japan*

Akio Miyao *Rice Genome Research Program, National Institute of Agrobiological Resources, 1-2 Kannondai 2-chome, Tsukuba, Ibaraki 305, Japan*

Simon P. Monard *Aaron Diamond AIDS Research Center, New York NY 10016, USA*

Richard Mott *Bioinformatics Department, SmithKline Beecham Pharmaceuticals, New Frontiers Science Park, Third Avenue, Harlow, Essex CM19 5AW, UK*

Michael A. North *Sequana Therapeutics, 11099 North Torrey Pines Road, Suite 160, La Jolla CA 92037, USA*

Gerald Nyakatura *Institute of Molecular Biotechnology, Department of Genome Analysis, Beutenbergstrasse 11, 07745 Jena, Germany*

Matthias Platzer *Institute of Molecular Biotechnology, Department of Genome Analysis, Beutenbergstrasse 11, 07745 Jena, Germany*

Guy Plunkett III *University of Wisconsin-Madison, Laboratory of Genetics, 445 Henry Mall, Madison WI 53706, USA*

Jaime Prilusky *Bioinformatics Unit, Department of Biological Services, Weizmann Institute of Science 76100, Rehovot, Israel*

Peter Richterich *Genome Therapeutics Corporation, 100 Beaver Street, Waltham MA 02154, USA*

André Rosenthal *Institute of Molecular Biotechnology, Department of Genome Analysis, Beutenbergstrasse 11, 07745 Jena, Germany*

Takuji Sasaki *Rice Genome Research Program, National Institute of Agrobiological Resources, 1-2 Kannondai 2-chome, Tsukuba, Ibaraki 305, Japan*

Robert D.C. Saunders *Department of Anatomy and Physiology, University of Dundee, Dundee DD1 4HN, UK*

Alan J. Schafer *Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK*

Leonard Schalkwyk *Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin, Germany*

Gabriele Senger *Institut für Humangenetik und Anthropologie, Kollegiengasse 10, D-07740 Jena, Germany*

Denise Sheer *Human Cytogenetics Laboratory, ICRF, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK*

Andrew N. Shelling *Department of Obstetrics and Gynaecology, National Women's Hospital, Claude Road, Epsom, Auckland, New Zealand*

Janet Shipley *Molecular Cytogenetics Laboratory, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK*

David L. Simmons *Cell Adhesion Laboratory, ICRF, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK*

John Sulston *The Sanger Centre, Hinxton Hall, Cambridge CB10 1SA, UK*

Karen Thomas *The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*

Rob B. van der Luijt *Department of Human Genetics, Faculty of Medicine, Utrecht University, Stratenum, De Uithof, Universiteitweg 100, 3584 CG Utrecht, The Netherlands*

Richard G. Vile *ICRF Laboratory of Cancer Gene Therapy, ICRF Oncology Unit, Hammersmith Hospital, DuCane Road, London W12 ONN, UK*

Sarah P. West *Northern Genetics Service, 19/20 Claremont Place, Newcastle upon Tyne NE2 4AA, UK*

Robert H. Waterston *Department of Genetics and Genome Sequencing Center, Washington University School of Medicine, St Louis, MO 63110, USA*

Kimiko Yamamoto *Rice Genome Research Program, Institute of Society for Techno-Innovation of Agriculture, Forestry and Fisheries, 446-1, Ippaizuka Kamiyokoba, Tsukuba, Ibaraki 305, Japan*

Masahiro Yano *Rice Genome Research Program, National Institute of Agrobiological Resources, 1-2 Kannondai 2-chome, Tsukuba, Ibaraki 305, Japan*



Preface



Genetics as the formal study of inheritance was founded as a field following the rediscovery of Mendel’s work at the beginning of this century. This led to the first revolution in our understanding of inheritance, namely of the basic mechanisms of gene transmission, of linkage and of interpretations in terms of the behaviour of chromosomes in meiosis. The second revolution came with the discovery of the Watson–Crick structure of DNA just over 40 years ago, which spelled out the chemical basis for the gene, and then its mode of action. Now, following the development of recombinant DNA technology and many other techniques that enable us to clone and sequence DNA with enormous speed and efficiency, we are entering a third revolutionary phase of genetic analysis as we approach the end of the century. Now is the time when whole genomes are being sequenced and the complete language of organisms is being deciphered.

It was just over 15 years ago that the potential for the complete analysis of the human genome began to be appreciated; it came to be realized that this would provide enormous power for the analysis of all normal biological functions, as well as for the analysis of the basis of essentially all human disease. Thus developed the Human Genome Project, and alongside it many other genome projects.

The rate of advance of the technology and the acquisition of new data could not, I believe, have

been predicted even by the wildest speculator. In 1986, I suggested that the project to catalogue and sequence all human genes and place them in their positions along the chromosomes be billed as ‘Project 2000’. That prediction we can now see will soon be realized.

Almost daily, new genes are discovered, while many exist and are waiting to be discovered in the databanks of genomic and, especially, partial cDNA sequences. The production and analysis of this extraordinary accumulation of information requires a wide variety of complex techniques; from approaches to the statistical problems of the analysis of complex human pedigrees, to the determination of DNA sequences. This *Handbook* provides an invaluable guide to the wide range of these techniques and is practical and usable. It has required an enormous effort on the part of the authors and, especially, the editors, to put together this most valuable companion and all ought to be congratulated on the achievement.

Only 5 years ago when we were organizing a new form of international Human Gene Mapping Workshop in London, it was hard to convince the pharmaceutical industry that they should be interested. Now, not only is there a huge and burgeoning biotechnology industry, but no major pharmaceutical company can afford any longer not to invest in a major way in genome analysis, and many are

accepting that this is where their future lies. The opportunities are enormous but the challenges now are to work with the genes and to understand their functions, and that may take perhaps another century or more to achieve. I am sure that this

Handbook will make an important contribution towards that end.

Walter Bodmer
ICRF, Laboratory Head

Introduction

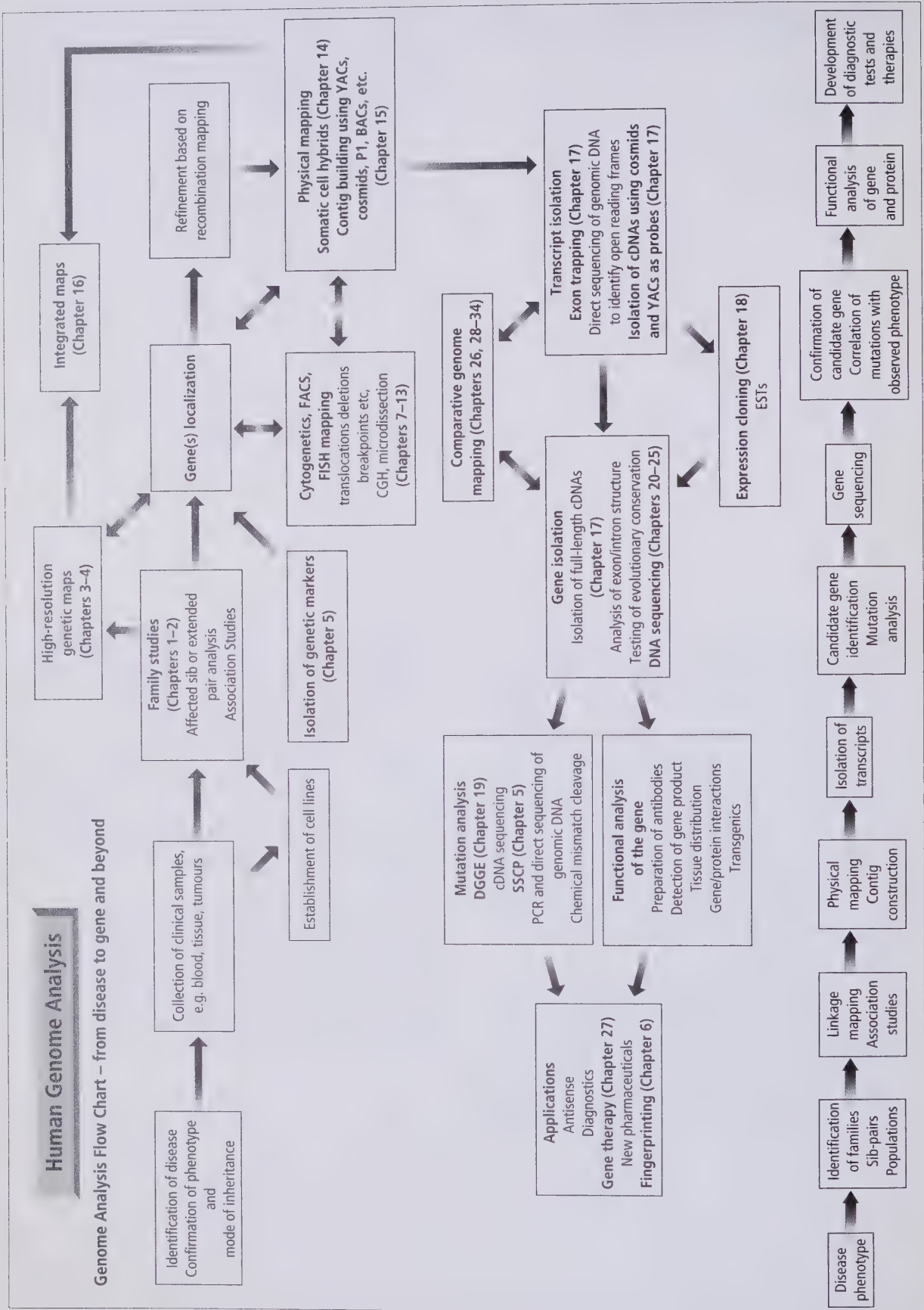
The *ICRF Handbook of Genome Analysis* is a combination of protocol manual and informational resource, drawing on the expertise of researchers at ICRF and elsewhere. It describes and evaluates a wide range of techniques pertinent to genome analysis. The first volume comprises a description and evaluation of strategies, techniques and protocols for use in the genetic and physical mapping of the human genome (Chapters 1–19). Genome analysis techniques are also used widely in the study and diagnosis of cancers and other diseases, and some of these applications are also covered. A glossary of abbreviations and acronyms is included at the end of Volume 2.

The second volume includes a comprehensive review section of approaches to DNA sequencing (Chapters 20–25) and reviews of progress in the analysis of the genomes of important model systems (Chapters 26–34). Organisms covered include the mouse, *Drosophila*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae* (the first eukaryote organism to have

its genome fully sequenced), *Escherichia coli*, *Arabidopsis thaliana* and rice. The second volume concludes with chapters on information resources and how to access them (Chapters 35–37) and appendices covering materials, preparation of blood samples, suppliers and other useful addresses, extensive tables of mapped human disease genes and mouse knockouts, and tables of chromosomal aberrations associated with cancer. An index to the complete handbook is included at the end of each volume.

One of the main driving forces behind the effort to map and sequence the human genome is the isolation and characterization of human disease genes. The figure on the following page shows the typical stages in such an enterprise and the relevant chapters in the *Handbook* that deal with the techniques involved.

Nigel K. Spurr





Section 1

Genetic mapping



Section 1 Introduction

Nigel K. Spurr

*SmithKline Beecham Pharmaceuticals, New Frontiers Science Park,
Harlow, Essex CM19 5AW, UK*

Genetic variation can be observed at many levels, most obviously as phenotypic variation—for example, in hair colour, fingerprints or protein variants. For many years, this allelic variation in proteins, which could be distinguished by electrophoresis, was the mainstay of human genetics; their genes were among the relatively few that could be mapped and used to test for linkage with suspected disease genes in families with inherited diseases.

Over the past 15 years, however, phenotypic variation as a means of linkage analysis has been replaced by direct genotyping, making use of polymorphisms detectable at the DNA level. The first such polymorphisms to be used were those within the recognition sites at which restriction endonucleases cleave DNA. More recently, with the introduction of the polymerase chain reaction (PCR) amplification system and the identification of short repeat sequences such as dinucleotide repeats, the whole area has been transformed. Only 6 years ago there were fewer than 500 markers detecting variation in humanoid DNA; currently there are over 10 000, the majority of which can be used in conjunction with PCR amplification. Chapter 5 (J. Armour) describes the various types of polymorphic markers now available and in Chapter 6 (J. Armour) their application in DNA fingerprinting for paternity testing and forensic medicine is also described.

These polymorphic markers are most commonly used in family studies to test for linkage to disease traits. Chapter 1 (S. West) details the problems associated with the collection of family material and the use of genetic markers in simple cases of Mendelian dominant and recessive traits. Chapter 2 (T. Bishop) expands on these applications by describing the problems and complexities of using markers in the study of complex diseases where the mode of inheritance is less clear. This area of genetics is going to be the most important in the next 10 years as we attempt to find the multiple genes involved in diseases such as asthma, diabetes and obesity.

The other major application of polymorphic markers is the construction of high-density genetic linkage maps of each chromosome. The increasing density of markers has in turn improved the speed and accuracy of mapping disease susceptibility genes in families. Once linkage is detected, in many cases there is now a wide range of closely flanking markers available to help confirm the results and to position markers on either side of the disease gene locus. The shorter the distance between the two flanking markers, the easier is the next stage of physical map construction which leads to the identification of the gene. Many genetic linkage maps showing the order and distance between markers have been published in recent years. The processes

involved in building such maps are described in Chapter 3 (S. Bryant) and in Chapter 4 (T.C. Matisse & A. Chakravarti) recent developments in the computer program MultiMap are described, which is able to eliminate many of the manual steps

and automate the process of map building. This type of linkage map has been used as the framework for the many integrated maps (see Chapter 16) now being produced.



Chapter 1

Genetic mapping:
family studies

Sarah P. West

Northern Genetics Service, 19/20 Claremont Place, Newcastle upon Tyne NE2 4AA, UK



1.1	Introduction and historical perspective, 6		
1.1.1	Meiotic segregation of genetic characters, 6	1.2.6.2	Calculation of a lod score for a two-generation family, 13
1.1.2	Non-independent segregation of genetic characters, 6	1.2.6.3	z_1 lod scores, 13
1.1.3	Measuring genetic distance, 6	1.2.6.4	z_2 lod scores, 14
1.1.4	Chromosome breakage and genetic recombination, 7	1.3	Collecting lod score data in practice, 17
1.1.5	Multipoint genetic maps, 7	1.3.1	Use of lod score tables, 17
1.1.6	Mapping functions, 7	1.3.2	Phase-known vs. phase-unknown linkage data, 17
1.2	Human gene mapping, 8	1.3.3	Maximizing the informativeness of linkage studies, 17
1.2.1	Human linkage studies, 8	1.3.3.1	Choice of markers, 17
1.2.2	Family studies, 8	1.3.3.2	Haplotype analysis, 21
1.2.2.1	Genetic markers, 8	1.4	Sex linkage, 22
1.2.2.2	Collecting family data, 9		
1.2.2.3	Phenotypes and genotypes, 9		Troubleshooting, 23
1.2.2.4	Penetrance, 10		Phenotype of interest does not segregate in a Mendelian pattern, 23
1.2.2.5	Phenocopies, 10		Family structure is not as recorded (i.e. non-paternity, adoption), 23
1.2.2.6	De novo mutations, 10		Linkage markers do not segregate in a Mendelian pattern, 23
1.2.3	Human linkage analysis, 10		Linkage markers segregate with disease in some families but not in others, 24
1.2.4	Standard error of a recombination fraction, 11		Haplotypes cannot be constructed, 24
1.2.5	Method of lod scores, 11		Distances between loci are incorrect, 24
1.2.5.1	Good evidence for linkage, 12		
1.2.5.2	Evidence against linkage, 12		References, 24
1.2.5.3	Sex differences in recombination fractions, 12		
1.2.6	Calculation of lod scores, 12		
1.2.6.1	Calculation of a lod score for a three-generation family, 13		

1.1 Introduction and historical perspective

1.1.1 Meiotic segregation of genetic characters

The foundations of modern genetics were laid in the 19th century, when Gregor Mendel studied the inheritance of pairs of discrete contrasting characters in the garden pea, *Pisum sativum*. From his observations, he deduced that each character of a pair was determined by ‘factors’, one of which was passed on at random to each offspring by a parent. The offspring thus had a new combination of characters, one from each parent. During gamete formation, each pair of characters was seen to segregate independently of the other pairs: this is Mendel’s law of independent segregation.

In 1903, from his studies on the segregation of homologous chromosomes at meiosis, Sutton [1] deduced that the chromosomes were the carriers of the Mendelian factors, or ‘genes’, as they had by then become known. Since the number of inherited characters exceeds the number of chromosome pairs, there must be several factors on each chromosome. The chromosomes were observed to segregate independently, which explained the independent segregation of factors carried on different chromosomes. But this could not explain the independent segregation of factors carried on the same chromosome, which would be expected to be inherited together. It was de Vries [2] in 1903 who pointed out that there could be exchanges of genes between the homologous chromosomes while they were paired at meiosis, and that this would account for the independent segregation of factors carried on the same chromosome.

1.1.2 Non-independent segregation of genetic characters

In 1905, Bateson, Saunders and Punnett [3] discover-

ed an important exception to the rule of independent segregation. Two pairs of contrasting characters in the sweet pea, *Lathyrus odoratus*, were observed to segregate in a non-independent manner. They observed many more parental types among the offspring than recombinant types, thus showing non-independent segregation or linkage. For example, suppose *A1* and *A2* represent the contrasting characters of one pair and *B1* and *B2* are the other pair. If an individual is known to have received *A1B1* from one parent and *A2B2* from the other parent (this is defined as knowing the ‘phase’ of the markers) then if independent segregation were occurring, this individual would be expected to produce equal numbers of *A1B1*, *A1B2*, *A2B1* and *A2B2* gametes. In the examples studied by Bateson and colleagues this was not the outcome; many more parental types (*A1B1* and *A2B2*), than recombinant (*A1B2* and *A2B1*) gametes were produced. In the nomenclature used today, *A1* and *A2* are called alleles; they are the alternative types observed at the locus *A*.

In 1911, T.H. Morgan [4] demonstrated another example of linkage, in the fruit fly *Drosophila*. The two segregating characters, white eye and miniature wing, were known to be carried on the X chromosome. More than 50% of the offspring of females segregating for these two characters were of the parental types; therefore, there was not independent segregation between these loci. Since some recombinant progeny were observed, and both these loci were known to lie on the X chromosome, this meant that genetic exchanges must have occurred between the two X chromosomes as had been predicted by de Vries. In 1912, Morgan and Cattell introduced the term ‘crossing-over’ to describe this exchange of characters between homologous chromosomes during gametogenesis [5].

1.1.3 Measuring genetic distance

This idea was developed further by Sturtevant [6], also investigating X-linked characters in *Drosophila*. He proposed that the genes are arranged in a linear order along the chromosome and that the frequency of crossing-over between two genes, which was observed to be almost constant for any particular pair of loci, measured the distance between them. Morgan had previously observed independent segregation between certain X-linked factors, and Sturtevant suggested that these loci were so far apart that a cross-over event between them was inevitable; they therefore appear to assort randomly at meiosis. This also explained, of course, how there could be more independently segregating genetic factors than chromosomes.

Family studies are used to:

- establish heritability of traits in families
- analyse the segregation of a trait in families
- establish a Mendelian pattern of inheritance for a trait — sex-linked or autosomal, dominant or recessive
- localize genes
- investigate candidate genes for their role in genetic disease
- construct genetic maps (see Chapter 3)
- identify risk genes in multifactorial diseases (see Chapter 2)

Applications box 1.1

The expression ‘recombination fraction’ was introduced to describe the proportion of the total number offspring that did not have a parental combination of alleles—that is, those that had a recombined pattern. The recombination fraction (RF) is defined by the expression:

$$RF = \frac{\text{number of recombinant offspring}}{\text{number of recombinant offspring} + \text{number of non-recombinant offspring}} \quad (1.1)$$

Recombination fractions could be used to measure the genetic distance between pairs of loci. The unit of measurement was called the Morgan (M) where one Morgan represents 100% recombination between the loci. Of course, 100% recombinant offspring were not observed, since even unlinked genes show only 50% recombination. In practice, the smaller units of centiMorgans (cM) are easier to use: 1 cM is equivalent to 1% recombinant offspring or a recombination fraction of 0.01.

1.1.4 Chromosome breakage and genetic recombination

The Chiasmatype Theory of Janssens [7] proposed a mechanism by which this genetic exchange could occur. It was known that during meiosis, the paternally inherited and maternally inherited chromosomes associated in their homologous pairs. Configurations looking like ‘bows’ were observed between pairs of chromatids derived from opposite homologues during this stage of meiosis. Janssens hypothesized that these bows, or chiasmata, represented the sites of chromosome breakage and reunion which allowed the exchange of genetic material between homologues. Direct evidence for the Chiasmatype Theory was not available since breakage and reunion events could not be seen through the light microscope. Furthermore, at the time it was generally considered that a chiasma was a region where the chromatids became tangled and a sufficiently accurate mechanism to carry out the precise breakage and reunion events was not thought possible.

It was not until the early 1930s that evidence to support the Chiasmatype Theory was forthcoming. Using two structural aberrations at opposite ends of a *Zea mays* chromosome, Creighton and McClintock demonstrated that these characters could segregate independently in meiosis, and therefore physical breakage and reunion of the chromatids must have occurred [8].

1.1.5 Multipoint genetic maps

The relationship between cross-over frequency and the genetically determined recombination fraction turned out not to be as simple as had been assumed originally. Sturtevant [6] showed that the greater the distance between two loci the greater the chance that more than one cross-over will occur between them. For instance, if the distance between two loci A and B is AB and between B and C is BC, the distance AC is found to be less than that predicted by AB and BC. From observations on three-point crosses, where three loci were segregating, it was clear that there could be double cross-overs between A and C. Thus the parental combinations of alleles are restored for the loci A and C and in such cases, the offspring would be scored as non-recombinants. The effect of double cross-overs is to reduce the apparent recombination fraction between loci that are more than a few centiMorgans apart, and the chance of a double cross-over event obviously increases as the genetic distance increases.

1.1.6 Mapping functions

In 1916, H. Muller [9] observed in *Drosophila* that the presence of one cross-over inhibits the occurrence of more cross-overs in the immediate vicinity. This phenomenon, called *interference*, is measured by the coefficient of coincidence, which is the ratio of the observed number of recombinants to the expected number of recombinants.

Several attempts have been made to define mathematically the relationship between the true map distances and the observed recombination fractions [10–12]. These mathematical functions—the *mapping functions*—describe curves that fit the data obtained from studying laboratory animals (*Drosophila* and the mouse). They all reach the conclusion that for map distances up to about 20 cM, the recombination fraction is an acceptably accurate estimate but that for greater distances, the relationship is unreliable.

The human mapping function was investigated by Sturt [13]. She points out that the distribution of chiasmata is affected by the requirement for an obligatory chiasma in each chromosome arm except the short arms of the acrocentric chromosomes (13, 14, 15, 21 and 22) [14]. Sturt demonstrates that there should be no interference between chiasmata across the centromere because obligate chiasmata on each side are independent. Second, the degree of interference varies according to the length of the chromosome arm rather than the map distance under consideration. Sturt’s curve indicates a moderately

high level of interference at map distances less than 30 cM. Rao *et al.* [15] derived the value of 0.35 for the coefficient of coincidence in metacentric chromosomes in male meiosis using the data of Hult  n [16] on the distribution of chiasmata at diakinesis. This indicates that the recombination fraction is a precise measurement of the map distance for intervals up to 35 cM. They suggest that for metacentric chromosomes in females, Kosambi’s estimate of the coincidence coefficient of 0.50 might be more reasonable. In acrocentric chromosomes, the coefficient of coincidence is less.

The curves shown in Fig.1.1 illustrate the relationship between map distance and recombination fraction according to Sturt across the centromere of a chromosome with arms of 1M in length, and according to Rao *et al.* for male metacentric chromosomes.

This concept of measuring the distance between loci by their recombination fractions and deriving a linear map of the genes is the essence of linkage investigations.

1.2 Human gene mapping

1.2.1 Human linkage studies

Linkage maps have been derived for many organisms, but, in the past, the constraints on carrying out this type of study on humans left the understanding of the human gene map many years behind those of other organisms. However, the important applications of a human gene map in both understanding

and managing genetic disease provided the impetus and justification for human linkage studies.

The problems of deriving a human linkage map are manifold. First is the long generation time, which means that an investigator is unlikely to be able to observe more than three generations within a pedigree. More commonly, only two generations are available for linkage investigations so that the amount of information derived from counting the proportion of recombinant offspring is very limited. Second is the problem of small family size, which means that data usually has to be pooled from several families in order to obtain a statistically reliable sample. Since it is not possible to set up desired crosses in humans as one can in animals, the investigator must search the population at large for the fortuitously informative families. The chance of finding suitable families depends on the gene frequencies in the population. Another problem frequently encountered in human genetic studies is non-paternity—the apparent father is not the natural father.

The intellectual and practical challenge posed by human gene mapping has provided incentives to circumvent many of these problems. Various statistical methods have been devised to maximize the informativeness of small families and two-generation families. Molecular genetics has provided the means to maximize the informativeness of genetic markers, which has, at the same time, enabled the detection of non-paternity.

1.2.2 Family studies

1.2.2.1 Genetic markers

In order for a locus to be suitable for linkage studies, it must show variation. The term polymorphism is used to describe a locus at which at least one in 50 unselected individuals has a variant allele; that is, the variant allele has a frequency greater than 0.01. Less polymorphic loci, such as the genes responsible for disease, may be used for linkage studies but it is then necessary to select these families from the population at large.

Typically, markers used for genetic studies are codominant, that is heterozygous individuals of the type A1A2 are distinguishable from homozygotes of the types A1A1 and A2A2. Most genetic diseases show dominant or recessive modes of inheritance. For example familial adenomatous polyposis coli (FAP or APC) is dominantly inherited, with affected individuals having one good copy of the APC gene and one defective copy. They then transmit the disease to half of all their offspring. With dominant diseases, affected individuals typically have an

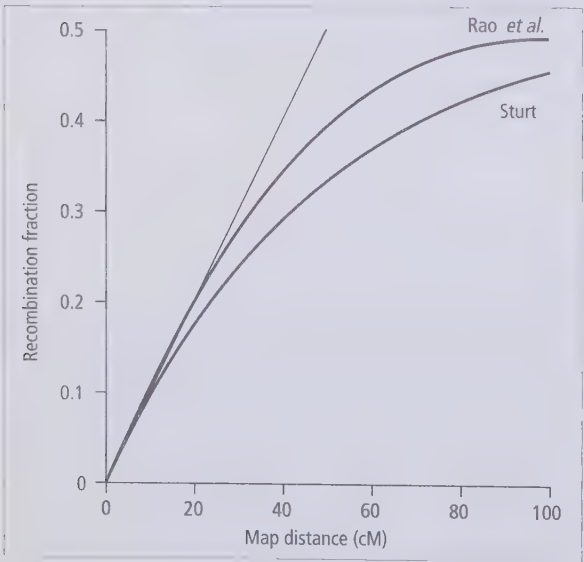


Fig.1.1 Relationship between the recombination fraction and map distance in cM according to Sturt [13] and Rao *et al.* [17].

affected parent. Exceptions to this occur when there is reduced penetrance (see later) or *de novo* mutations.

With recessive conditions such as cystic fibrosis (CF), affected individuals have defects in both copies of the responsible gene. The parents of individuals affected by recessive diseases are usually unaffected because they are heterozygous carriers with one normal copy and one defective copy of the gene. The normal allele is sufficient to provide normal function and so no disease phenotype is observed. Recessive traits are characterized by clustering in sibships but are seldom seen in successive generations of pedigrees. For this reason they have proved a far more challenging problem than dominant traits for linkage analysts.

1.2.2.2 Collecting family data

By convention, family structure is illustrated by a family tree or pedigree (see Figs 1.3–1.5). These charts are a convenient way of describing diagrammatically the relationships within a family. The symbols most frequently used are shown in Fig. 1.2.

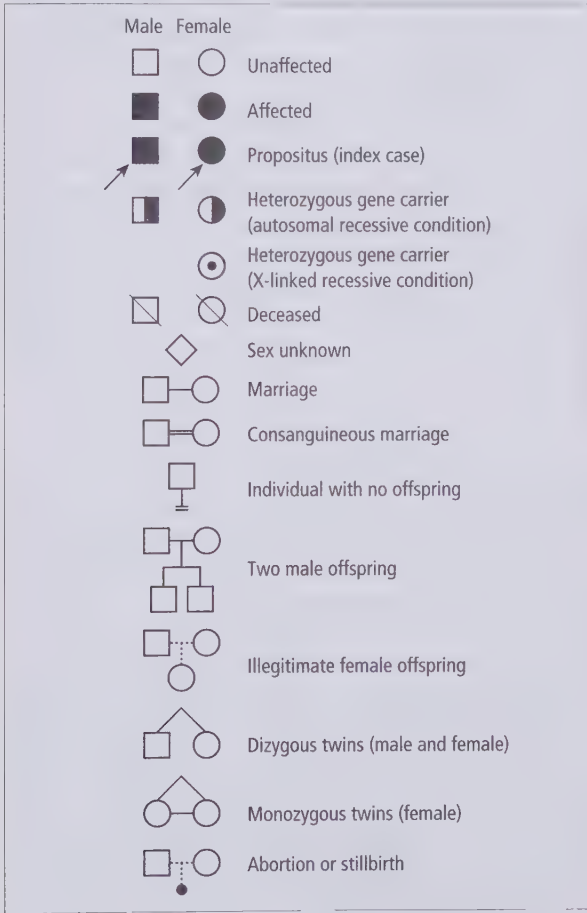


Fig. 1.2 Frequently used symbols in pedigree drawing for human genetics.

Members of the same generation (siblings and spouses) are placed at the same horizontal level, with younger generations below and older ones above. Generations are numbered from the top downwards on the left-hand side by convention using Roman numerals. Individuals are numbered from left to right within a generation with the offspring of each marriage placed in birth order with the eldest on the left.

It is essential to record on the pedigree the full names (including names before marriage, etc.) of all family members and their dates of birth, even those who may not be involved with the study, in order to prevent confusion later. Questions relating to other family members must be asked if the family is being investigated for an inherited disease. Information about the causes and dates of death of relatives, information about infant deaths, stillbirths and abortions, and also about previous marriages and consanguinity may be very important but will need to be sought with discretion. Since the families are the most valuable and irreplaceable resource for human genetics it is paramount that they are treated with the utmost sensitivity and caused the minimum inconvenience possible.

It may be necessary to seek confirmation of diagnosis of a genetic disorder from clinical records, since a watertight diagnosis underpins any attempt to genetically map a causative gene. The effort made to record these data will pay dividends at later stages of the study when things inevitably turn out to be more complex than originally estimated.

1.2.2.3 Phenotypes and genotypes

In their enthusiasm for a genetic explanation for the phenomenon being considered, it is not unknown for geneticists to overlook the influence of epigenetic factors in determining human characteristics. Even in the cases of apparently monogenic disorders the phenotype of an affected individual results from the combined effects of their genotype and environment. For example, two cystic fibrosis patients both homozygous for the *CFTR* ΔF508 mutation, the most common in the UK population, will exhibit different combinations of the symptoms typical of CF. This is due to the influence of different environmental factors and of their different genetic backgrounds—that is, other genes in their genomes, each with very subtle effects. The variation in phenotype between individuals with the same genetic condition is described as variable expressivity. In FAP, variable expressivity is very marked, with some patients having fewer than 100 colonic adenomatous polyps in their fourth decade or so of life. Other patients may have thousands of polyps by their teenage

years, together with a range of extracolonic manifestations including upper gastrointestinal tract polyps, epidermoid cysts, osteoma, desmoid tumours and a high risk of cancers in other organs.

1.2.2.4 Penetrance

An extreme example of variable expressivity is when known gene carriers do not express any symptoms of the disease at all. This is described as reduced penetrance of the disease gene, or incomplete penetrance. A few examples of reduced penetrance are recorded for FAP, where transmission of the disease from an affected grandparent to an affected grandchild has involved an apparently unaffected parent. The frequency of these events can be estimated, and describes the percentage penetrance of the disease gene. The penetrance of FAP mutations is around 95% by age 50. Hereditary non-polyposis colon cancer (HNPCC) is a dominantly inherited disorder predisposing disease gene carriers to colorectal cancer at a younger age than the general population. HNPCC is due to defects in genes encoding components of the DNA mismatch repair mechanism which corrects errors occurring during DNA synthesis. In this disorder, lifelong penetrance is estimated to be only 80%, meaning that a mutant gene carrier has an 80% risk of developing colorectal cancer or an associated cancer during his or her lifetime. Therefore it is possible that an apparently unaffected individual is a non-manifesting gene carrier and would appear in linkage studies to be a recombinant between the disease locus and a closely linked marker.

To reduce errors introduced into linkage studies by the misclassification of non-penetrant individuals, age of onset or penetrance curves are used. These relate the age of the unaffected, at-risk individual with his or her risk of manifesting the disease, and weight the data interpretation for that individual accordingly. For example, an unaffected at-risk 20-year-old from an HNPCC family has a greater risk of carrying the disease gene than an at-risk family member who has reached the age of 80 with no detectable symptoms.

1.2.2.5 Phenocopies

In considering conditions such as HNPCC, the phenotype of colon cancer is also a common occurrence in the general population as a sporadic event not due to defective mismatch repair genes. These sporadic colorectal cancers are phenocopies of the genetically determined phenotype.

Care is needed when selecting families for linkage studies in such situations where non-genetic phenocopies of the genetic trait are common. For example,

for a family to be classified as an HNPCC kinship it has to fulfil a list of criteria known as Amsterdam criteria, such as three first-degree relatives affected with colorectal cancer, one of which has been diagnosed at an early age [17]. This reduces the chance of including families in which there is a familial clustering of colorectal cancer phenocopies.

1.2.2.6 De novo mutations

Normal patterns of inheritance in families are disturbed when high rates of mutation occur in the genes under investigation. It is estimated that one in three Duchenne muscular dystrophy (DMD) patients is the result of a spontaneous mutation event with no previous family history of the disease [18]. Many other disease genes also show unexpectedly high *de novo* mutation rates; approximately one-fifth of probands in FAP kindreds are the result of new mutations.

The mechanisms underlying elevated mutation rates are many and various. For example, the exceptionally large size—approximately 2 Mb—of the genomic region encoding the dystrophin gene (the affected gene in DMD), is believed to make it an exceptionally good target for random mutational events. For FAP, short repeated sequences within the *APC* gene are believed to promote DNA polymerase slippage during DNA replication and hence introduce the small insertion and deletion mutations characteristic of this gene. The stretches of tandemly repeated trinucleotide sequences which undergo expansion to large alleles in diseases such as myotonic dystrophy, Huntington's disease and fragile X syndrome, are believed to be inherently unstable.

1.2.3 Human linkage analysis

Human linkage analyses require families where one parent is heterozygous at the two loci to be tested, that is having distinguishable alleles, and where the segregation of alleles at these loci can be observed in the offspring. In the simplest situation, it is possible to determine which alleles at the two loci are associated on the same chromosome from the grandparents in the pedigree. This valuable information is often described as the 'phase' of the double heterozygote. This is illustrated in Fig. 1.3. The doubly heterozygous parent II.1 has received the alleles *A1* and *B1* from his father and *A2* and *B2* from his mother. Therefore his 'phase' is defined because it is evident that *A1* and *B1* must lie on the same chromosome, with *A2* and *B2* on the other chromosome.

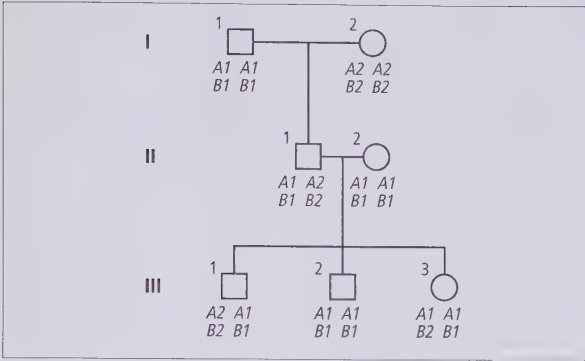


Fig. 1.3 Segregation of alleles *A1* and *A2* at the locus *A* and *B1* and *B2* at the locus *B* in a family where phase in the doubly heterozygous parent (II.1) can be determined from the grandparents.

The children of II.1 and II.2 have all inherited *A1B1* from their mother, since she is homozygous at these loci; that is, both of her chromosomes carry the same allele at these two loci and are therefore indistinguishable. II.1 has transmitted to his first son, III.1, his maternally derived chromosome bearing the alleles *A2* and *B2*. To his second son III.2, he has transmitted the paternally derived *A1B1* chromosome. His daughter, III.3, has inherited the alleles *A1* and *B2* from her father. These alleles came from different grandparents and so a genetic cross-over must have occurred in II.1 during the meiotic events resulting in the gamete which formed III.3. Therefore, of the three children of II.1, one is a recombinant between the loci *A* and *B* and so the recombination fraction (RF) can be defined by the expression:

$$\begin{aligned} \text{RF} &= \frac{\text{number of recombinants}}{\text{total number of recombinants} + \text{non-recombinants}} \\ &= \frac{1}{1 + 2} \\ &= 1/3 \end{aligned} \tag{1.2}$$

Information can be pooled from many families to make a direct estimate of the recombination fraction, defined by the expression:

$$\text{RF} = \frac{\text{total number of recombinants}}{\text{total number of recombinants} + \text{non-recombinants}} \tag{1.3}$$

1.2.4 Standard error of a recombination fraction

Since the recombination fraction merely estimates

the absolute value of the genetic distance then an assessment of its accuracy should be made, this is the standard error (s.e.). The standard error is calculated as that for a binomial distribution because recombination events are discrete events. The expansion of the binomial distribution is given by $(p+q)^n$, where $p = (1-q)$, which is the chance of observing a recombinant. So, for example, if a recombinants are observed in a sample of n offspring, then a/n estimates the recombination fraction p . In reasonably large samples—say, where $n > 30$ —the binomial distribution fraction and its error are described by:

$$a/n \pm \sqrt{\frac{a/n (1 - a/n)}{n}} \tag{1.4}$$

such that the error of the recombination fraction is:

$$\sqrt{\frac{\text{recombination fraction} \times \text{non-recombination fraction}}{\text{total number of offspring}}} \tag{1.5}$$

In practice, many investigators prefer to use ‘confidence limits’ to describe their results. To calculate these confidence limits requires that the mean value has a normal distribution, this assumption is reasonable for large samples where $n > 30$. The 95% confidence interval is most commonly used; this defines the range within which experimental estimates of the true mean are correct in 95% of cases. The 5% of cases which are outside the confidence limits lie more than 1.96 standard deviations (s.d.) from the mean. For example, if 25 recombinants were observed out of 100 offspring, then the estimate of the recombination fraction and its standard error are:

$$\begin{aligned} \frac{25}{100} \pm \sqrt{\frac{0.25 \times 0.75}{100}} \\ = 0.25 \pm 0.043 \end{aligned} \tag{1.6}$$

and the 95% confidence interval is 0.166–0.334.

1.2.5 Method of lod scores

Unfortunately, three-generation pedigrees are seldom available for genetic linkage studies, and so direct estimation of the recombination fraction is not possible. Several methods for indirect estimation have been developed, but it is the sequential method of lod scores which is most commonly used [19]. It is applied according to Maynard-Smith *et al.* [20], often with the computing assistance of software packages such as LIPED [21] and LINKAGE [22] (see Chapter

3 for an example of the use of LINKAGE). This method has the advantages that information from phase-known and phase-unknown linkage data can be combined, and data generated by different research groups can be pooled even if the raw pedigree data are not available.

The lod score method compares the probability of obtaining the offspring observed from a given mating if the two loci being considered are linked at a defined recombination fraction (θ) with the probability of obtaining these offspring if the loci are unlinked, that is, the recombination fraction (θ) = 0.50. A range of values for θ are generally used in the calculations, for example, 0.01 for close linkage, 0.05, 0.10, 0.20, 0.30 and up to 0.40 for loose linkage.

In order to make pooling data from many small sibships more convenient, the logarithm to base 10 of the ratio of the odds on linkage is taken, hence the name 'lod' from 'log odds ratio'. Since each offspring, except monozygotic twins, is the product of meiotic events which are independent of the events giving rise to siblings, then in calculating the overall probability ratio for a family, the contributions of each child should be multiplied together to reach the total. This enables the scores from several families to be added together and so simplifies collecting a total estimate of the odds on linkage.

A lod score, z , may be defined by the expression:

$$z = \log_{10} \frac{\text{probability of observing this family if loci linked with recombination fraction of } \theta}{\text{probability of observing this family if loci not linked, i.e. } \theta = 0.50} \quad (1.7)$$

The lod score for $\theta = 0.50$ is always zero since the probability ratio is one. Positive lod scores point towards linkage between the two loci, and negative lod scores decrease the chance of linkage. The value of θ where the lod score is largest is the maximum likelihood estimate of the recombination fraction between the two loci.

1.2.5.1 Good evidence for linkage

Many linkage workers adopt the convention when dealing with autosomal loci that when the peak lod score exceeds +3.0, that is when the maximum antilod exceeds 1000, then there is convincing evidence for linkage. The reason for using this value seems rather obscure and according to Smith and Sturt [23], a better criterion on which to decide whether or not two loci are linked is calculated from

the average height (\bar{H}) of the antilod curve, between values of θ from 0 to 0.5.

Before anything else is known about the relationship between a pair of autosomal loci, the prior probability against their linkage is 21 : 22 (that is, the probability that they are on different chromosomes) and the prior probability that they are linked is 1 : 22. Thus, the prior odds against linkage are 21 : 1. In practice, the value of 19 is used for the prior odds against linkage because this takes into account the differences in chromosome lengths.

Combining these assumptions with the observed data, where \bar{H} gives the average value for the odds in favour of linkage then the total odds in favour of linkage are:

$$\frac{\bar{H}}{\bar{H} + 19} \quad (1.8)$$

Thus, if $\bar{H} > 20$, then the odds are in favour of linkage.

It is very difficult to determine how \bar{H} is in general related to the peak of the likelihood curve. From considering typical cases, Smith and Sturt [23] suggest that if the peak value of the likelihood curve is 1000 — that is, a maximum lod score of +3.0 — then this implies a linkage probability of about 90%.

1.2.5.2 Evidence against linkage

A lod score value of -2 or less (odds of 100 : 1 against linkage) is frequently accepted as indicating that the two loci are not linked at that particular recombination fraction. The possibility that the loci may be linked at a different recombination fraction must not be overlooked.

1.2.5.3 Sex differences in recombination fractions

Generally, recombination is seen to be more frequent in female meioses than in male meioses [24,25]. For this reason, it is easier to demonstrate linkage in males, and when collecting lod score data, it is helpful to separate them by the sex of the double heterozygous parent.

1.2.6 Calculation of lod scores

The Greek symbols θ (theta) and ψ (psi) are used in deriving lod score expressions where θ is the recombination fraction between the two markers and $\psi = 1 - \theta$ such that ψ is the 'non-recombination fraction'. The subscripts 'm' and 'f' may be used to indicate whether θ or ψ pertain to male or female meioses; for example, θ_m is the male recombination fraction.

1.2.6.1 Calculation of a lod score for a three-generation family

For example, in the pedigree illustrated in Fig. 1.3, the relative probability of linkage between A and B for the first child III.1 can be calculated as follows:

$$\frac{\text{chance of receiving A2 from II.1} \times \text{chance of receiving B2 from II.1 given A2 if loci are linked}}{\text{chance of receiving A2 from II.1} \times \text{chance of receiving B2 from II.1 given A2 if loci are not linked}} = \frac{\frac{1}{2} \times \psi}{\frac{1}{2} \times \frac{1}{2}} = 2\psi \tag{1.9}$$

For the second child, III.2, the relative probability of linkage is:

$$\frac{\text{chance of A1 from II.1} \times \text{chance of B1 from II.1 given A1 if loci are linked}}{\text{chance of A1 from II.1} \times \text{chance of B1 from II.1 given A1 if loci are not linked}} = \frac{\frac{1}{2} \times \psi}{\frac{1}{2} \times \frac{1}{2}} = 2\psi \tag{1.10}$$

For the third child, III.3, the relative probability of linkage is:

$$\frac{\text{chance of A1 from II.1} \times \text{chance of B2 from II.1 given A1 if loci are linked}}{\text{chance of A1 from II.1} \times \text{chance of B2 from II.1 given A1 if loci are not linked}} = \frac{\frac{1}{2} \times \theta}{\frac{1}{2} \times \frac{1}{2}} = 2\theta \tag{1.11}$$

Since each child is the product of independent meiotic events, then the overall lod score, *z*, for this family is the log₁₀ of the product of these probabilities. That is:

$$z = \log_{10} (2^3 \cdot \psi^2 \cdot \theta)$$

In general, the lod score for a phase-known family is:

$$z = \log_{10} (2^s \cdot \psi^a \cdot \theta^b)$$

where

a = number of non-recombinant offspring

b = number of recombinant offspring

s = a + b = total number offspring scored

1.2.6.2 Calculation of a lod score for a two-generation family

In two-generation families and other situations where there is no information on the phase of the double heterozygote, it is assumed that both phases are equally likely. In order to calculate a lod score, the probability of observing such a family, with and without linkage, must be computed separately for each phase. Then the final lod score is obtained from the mean of these two probabilities.

1.2.6.3 *z*₁ lod scores

*z*₁ lod scores are the most commonly used lod scores in phase-unknown situations, since these are appropriate where there is no recessivity. This occurs when dealing with codominant variants at the molecular genetic level and dominantly inherited diseases where, because they are rare in the population, the chance of observing a homozygous patient is negligible under normal circumstances.

The pedigree of another family segregating for alleles at the loci A and B is illustrated in Fig. 1.4. Since this is a two-generation family, there is no information concerning the phase of alleles at the two loci, and there are two equally likely possibilities for the phase in the double heterozygote I.1:

$$\text{phase 1 } \frac{A1 \ B1}{A2 \ B2} \qquad \text{phase 2 } \frac{A1 \ B2}{A2 \ B1} \tag{1.12}$$

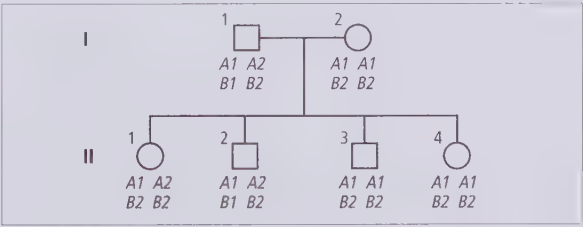


Fig. 1.4 Segregation of alleles A1 and A2 at the locus A and B1 and B2 at the locus B in a family where the phase of the doubly heterozygous parent (I.1) is unknown.

If phase 1 is considered first, then the odds on linkage are calculated for the first child, II.1, using:

$$\frac{\text{chance of receiving A2 from I.1} \times \text{chance of receiving B2 from I.1 given A2 and if loci are linked}}{\text{chance of receiving A2 from I.1} \times \text{chance of receiving B2 from I.1 given A2 and if loci are not linked}} \tag{1.13}$$

$$= \frac{\frac{1}{2} \times \psi}{\frac{1}{2} \times \frac{1}{2}} \tag{1.14}$$

Similarly, the relative probabilities for linkage if I.1 is phase 1 are calculated for the three other offspring. The total odds on linkage for this family if I.1 is phase 1 are:

$$2^4 \times \psi^1 \times \theta^3$$

Then phase 2, which is equally likely, must be considered. The odds on linkage from this family, given phase 2 for I.1, are:

$$2^4 \times \psi^3 \times \theta^1$$

The overall odds on linkage contributed by this family are derived by taking the mean of the odds from both of these phase hypotheses:

$$\frac{2^4 \times \psi \times \theta^3 + 2^4 \times \psi^3 \times \theta}{2} \tag{1.15}$$

Therefore, the total lod score, z_1 ,

$$= \log_{10} 2^3 (\psi \cdot \theta^3 + \psi^3 \cdot \theta)$$

The general expression for a z_1 lod score may be written as:

$$z_1 = \log_{10} 2^{s-1} (\psi \cdot \theta^b + \psi^b \cdot \theta^a)$$

where
a = number offspring which are non-recombinants if phase 1
b = number offspring which are non-recombinants if phase 2
s = a + b = total number offspring scored

The shorthand expression z_1 a:b for the z_1 lod score observed for a particular family is often used for convenience. Since the odds on linkage are the same for a z_1 a:b score as a z_1 b:a score, it is customary to write the larger number first. Therefore the shorthand expression for the lod score in this example is z_1 3:1.

1.2.6.4 z_2 lod scores

A more complicated system of scoring is required to deal with recessive characters in certain situations. For instance, in dealing with blood group markers and in the case of recessive diseases, the parents are shown to be heterozygous by the occurrence of a recessive homozygous offspring, such as a child with cystic fibrosis of phenotypically normal parents where mutation analysis has not been performed. In such cases, a z_2 lod score is appropriate.

The pedigree of a family in which a recessive disease, such as cystic fibrosis, occurs is shown in

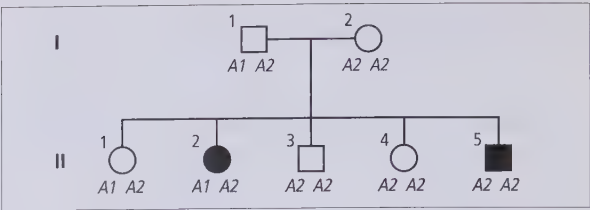


Fig. 1.5 Segregation of alleles in a family segregating for a recessive condition such as cystic fibrosis.

Fig. 1.5. In this family, alleles $A1$ and $A2$ at the locus A are seen to segregate and the parents I.1 and I.2 are assumed to be heterozygous carriers for a recessive genetic disorder because they have two affected children II.2 and II.5. This family could be scored in two equally valid ways. The simpler method is to calculate a z_1 lod score on the two recessive homozygotes since it is clear that each parent has contributed an affected allele, d , for the disease locus D . Therefore, I.1 has transmitted d and $A1$ to II.2 and d and $A2$ to II.5 and so the relative likelihood ratio is $2(\psi\theta + \theta\psi)$ and so the z_1 lod score is $\log_{10} [2(2\theta\psi)]$.

This method, of course, cannot use the information contributed by the unaffected offspring who may be homozygous, DD , or heterozygous carriers, Dd , with the affected allele, d , having been transmitted by either father or mother. The z_2 lod score method enables information from these unaffected children to be included. There are two equally likely possibilities for phase in the doubly heterozygous father I.1 which are:

$$\text{phase 1 } \frac{A1 \ D}{A2 \ d} \quad \text{or} \quad \text{phase 2 } \frac{A2 \ D}{A1 \ d} \tag{1.16}$$

As with a z_1 lod score, the odds on linkage provided by this family are the average of the odds for each phase, but in this case, the contribution of the mother, I.2, at the D locus also needs to be taken into account.

The phase of the mother, I.2, must be:

$$\frac{A2 \ D}{A2 \ d} \tag{1.17}$$

To calculate the relative odds on linkage if father, I.1, is phase 1:

The first child, II.1, is $A1A2$ unaffected. It is clear that she must have received the $A1$ allele from her father, but, whether she received a normal allele, D , for the disease locus from father, mother or both is unknown. Therefore, the probability of observing this child is the sum of the probabilities of the two mutually exclusive possibilities that her father contributed (a) the D allele or (b) the d allele. Table 1.1 shows how to calculate the probability if either

Table 1.1 Calculation of the relative probability for child II.1 if phase 1 is assumed for I.1.

	If loci are linked	If loci are unlinked
Phase 1 and <i>D</i> from I.1		
Chance of receiving <i>A1</i> from I.1	$\frac{1}{2}$	$\frac{1}{2}$
Chance of receiving <i>D</i> from I.1 given <i>A1</i>	ψ	$\frac{1}{2}$
Chance of receiving <i>A2</i> from I.2	1	1
Chance of receiving <i>D</i> or <i>d</i> from I.2 given <i>A2</i>	1	1
Phase 1 and <i>d</i> from I.1		
Chance of receiving <i>A1</i> from I.1	$\frac{1}{2}$	$\frac{1}{2}$
Chance of receiving <i>d</i> from I.1 given <i>A1</i>	θ	$\frac{1}{2}$
Chance of receiving <i>A2</i> from I.2	1	1
Chance of receiving <i>D</i> from I.2 given <i>A2</i>	$\frac{1}{2}$	$\frac{1}{2}$

situation applies. Therefore, the relative probability of observing this offspring:

$$\begin{aligned} &= \frac{\frac{1}{2} \cdot \psi \cdot 1 \cdot 1 + \frac{1}{2} \cdot \theta \cdot 1 \cdot \frac{1}{2}}{\frac{1}{2} \cdot \frac{1}{2} \cdot 1 \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} \cdot 1 \cdot \frac{1}{2}} \\ &= \frac{2\psi + \theta}{1 + \frac{1}{2}} \\ &= \frac{1 + \psi}{3/2} \quad (\text{since } \psi + \theta = 1) \end{aligned} \tag{1.18}$$

Let a = number of *A1A2* unaffected offspring, then the contribution of all the *A1A2* unaffected offspring to the odds ratio:

$$= \frac{(1 + \psi)^a}{(3/2)} \tag{1.19}$$

The second child, II.2, is *A1A2* affected so she must have received *A1* and *d* from I.1. The relative probability of observing this offspring if phase 1 is shown in Table 1.2. Therefore, the relative probability of observing this offspring:

$$\begin{aligned} &= \frac{\frac{1}{2} \cdot \theta \cdot 1 \cdot \frac{1}{2}}{\frac{1}{2} \cdot \frac{1}{2} \cdot 1 \cdot \frac{1}{2}} \\ &= 2\theta \end{aligned} \tag{1.20}$$

Let b = number of *A1A2* affected offspring, then the contribution of all the *A2A2* affected offspring to the total odds ratio on linkage for this family if phase 1 is (2 θ)^b.

Table 1.2 Calculation of the relative probability for child II.2 if phase 1 is assumed for I.1.

	If loci are linked	If loci are unlinked
Chance of receiving <i>A1</i> from I.1	$\frac{1}{2}$	$\frac{1}{2}$
Chance of receiving <i>d</i> from I.1 given <i>A1</i>	θ	$\frac{1}{2}$
Chance of receiving <i>A2</i> from I.2	1	1
Chance of receiving <i>d</i> from I.2 given <i>A2</i>	$\frac{1}{2}$	$\frac{1}{2}$

The third child, II.3, is *A2A2* unaffected. She must have received *A2* from her father, but again, it is unclear whether she has received *D* from father, mother or both. The probabilities, for each situation, that I.1 contributed (a) *D* or (b) *d* must be calculated (Table 1.3). Therefore, the relative probability of observing this offspring:

$$\begin{aligned} &= \frac{\frac{1}{2} \cdot \theta \cdot 1 \cdot 1 + \frac{1}{2} \cdot \psi \cdot 1 \cdot \frac{1}{2}}{\frac{1}{2} \cdot \frac{1}{2} \cdot 1 \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} \cdot 1 \cdot \frac{1}{2}} \\ &= \frac{2\theta + \psi}{1 + \frac{1}{2}} \\ &= \frac{1 + \theta}{3/2} \end{aligned} \tag{1.21}$$

Let c = number of *A2A2* unaffected offspring, then the contribution of all the *A2A2* unaffected offspring to the total odds ratio for this family if:

$$= \frac{(1 + \theta)^c}{(3/2)} \tag{1.22}$$

The fourth child, II.4, has the same genotype as the third and so for this family, c = 2.

The fifth child, II.5, is *A2A2* affected and so has received *A2* and *d* from I.1. Calculation of the phase 1 relative probability of observing this offspring is shown in Table 1.4. Therefore, the relative probability of observing this offspring:

	If loci are linked	If loci are unlinked
Phase 1 and D from I.1		
Chance of A2 from I.1	$\frac{1}{2}$	$\frac{1}{2}$
Chance of D from I.1 given A2	0	$\frac{1}{2}$
Chance of A2 from I.2	1	1
Chance of D or d from I.2 given A2	1	1
Phase 1 and d from I.1		
Chance of A2 from I.1	$\frac{1}{2}$	$\frac{1}{2}$
Chance of d from I.1 given A2	ψ	$\frac{1}{2}$
Chance of A2 from I.2	1	1
Chance of D from I.2 given A2	$\frac{1}{2}$	$\frac{1}{2}$

Table 1.3 Calculation of the relative probability for child II.3 if phase 1 is assumed for I.1.

	If loci are linked	If loci are unlinked
Chance of A2 from I.1	$\frac{1}{2}$	$\frac{1}{2}$
Chance of d from I.1 given A2	ψ	$\frac{1}{2}$
Chance of A2 from I.2	1	1
Chance of D from I.2 given A2	$\frac{1}{2}$	$\frac{1}{2}$

Table 1.4 Calculation of the relative probability for child II.5 if phase 1 is assumed for I.1.

$$\begin{aligned} &= \frac{\frac{1}{2} \cdot \psi \cdot 1 \cdot \frac{1}{2}}{\frac{1}{2} \cdot \frac{1}{2} \cdot 1 \cdot \frac{1}{2}} \\ &= 2\psi \end{aligned} \tag{1.23}$$

Let d=number of A2A2 affected offspring, then the contribution of all the A2A2 affected offspring to the total odds ratio on linkage for this family if phase 1 is (2ψ)^d.

The expression for the relative probability of linkage if phase 1 for the whole family is the product of the four expressions:

$$\frac{(1 + \psi)^a}{3/2} \cdot \frac{(2\theta)^b}{3/2} \cdot \frac{(1 + \theta)^c}{3/2} \cdot (2\psi)^d \tag{1.24}$$

If
s=total number offspring scored
= a + b + c + d
then the relative probability of linkage if phase (1) can be expressed as:

$$\frac{2^s}{3^{(a+c)}} \left((1 + \psi)^a \cdot \theta^b \cdot (1 + \theta)^c \cdot \psi^d \right) \tag{1.25}$$

Conversely, if the phase in I.1 is phase 2, then the expression for the relative probability of linkage is:

$$\frac{2^s}{3^{(a+c)}} \left((1 + \theta)^a \cdot \psi^b \cdot (1 + \psi)^c \cdot \theta^d \right) \tag{1.26}$$

The total z₂ lod score expression for this family is derived from the log₁₀ of the average of these two relative probabilities of linkage:

$$z_2 = \log_{10} \frac{2^4}{3^3} \frac{[(1 + \psi)\theta(1 + \theta)^2 \psi + (1 + \theta)\psi(1 + \psi)^2 \theta]}{\tag{1.27}}$$

The general expression for a z₂ lod score is:

$$\frac{2^{(s-1)}}{3^{(a+c)}} \frac{[(1 + \psi)^a \theta^b (1 + \theta)^c \psi^d + (1 + \theta)^a \psi^b (1 + \psi)^c \theta^d]}{\tag{1.28}}$$

- where:
- a=number of dominant phenotype (unaffected) offspring who have inherited the first allele (A1) at the marker locus
 - b=number of recessive phenotype (affected) offspring who have inherited the first allele (A1) at the marker locus
 - c=number of dominant phenotype (unaffected) offspring who have inherited the second allele (A2) at the marker locus
 - d=number of recessive phenotype (affected) offspring who have inherited the second allele (A2) at the marker locus
 - s=total number offspring scored
= a + b + c + d

The shorthand expression z₂ a : b : c : d for the z₂ lod score observed for a particular family is used for convenience.

1.3 Collecting lod score data in practice

In practice, these calculations are not carried out for each family scored. Most workers use computer programs for linkage analyses such as LIPED [21] and LINKAGE [22], which utilize pedigree data to generate lod score tables. Pedigree data can be entered directly into these programs but the use of data management systems simplifies this error-prone stage and facilitates double-checking of data and reformatting data for different applications. Two systems in common use are LINKSYS [26] and CYRILLIC. However, it is important to be able to check the output from computer analyses, and lod score tables are used for this purpose.

1.3.1 Use of lod score tables

Despite the availability of these computer programs it is none the less useful to be able to check their output or make estimates of linkage data by carrying out hand calculations. These involve the use of tables of lod scores computed for various values of θ for each possible class of family, such as given in Table 44 of Maynard-Smith *et al.* [20]. Similar tables for some useful θ values are given in Tables 1.5–1.7.

The lod scores for phase-known direct counting of recombinants and non-recombinants are given in Table 94 of Race and Sanger [27] and in an abbreviated form in Table 1.5. For example, the lod scores for the family shown in Fig. 1.2, with two non-recombinant (NR) and one recombinant (Rec) offspring (i.e. 2NR : 1Rec) at recombination fractions of 0.01, 0.05, 0.10 and 0.30 are -1.106 , -0.442 , -0.189 and $+0.070$, respectively.

For phase-unknown codominant markers where z_1 scores are appropriate, Table 1.6 is used. The table is entered for the corresponding z_1 score by counting the numbers of children in each of classes a and b, by convention writing the larger number first. The z_1 example calculated for the pedigree in Fig. 1.4 (i.e. z_1 3 : 1) at recombination frequencies of 0.01, 0.05, 0.10 and 0.30 yields the following lod scores: -1.110 , -0.464 , -0.229 and -0.011 , respectively.

In a similar manner, z_2 scores are obtained from Table 46 in Maynard-Smith *et al.* [20]. An abbreviated form of this is given in Table 1.7 for some useful values of θ . The appropriate lod scores are found by counting the numbers of children in each of classes a, b, c and d and then using these values to enter the table. If the corresponding z_2 class does not appear in the table, remember that a : b : c : d can be rewritten c : d : a : b without confounding the score. Thus the z_2

1 : 1 : 2 : 1 worked example at recombination frequencies of 0.01, 0.05, 0.10 and 0.30 gives the following lod scores: -1.451 , -0.762 , -0.467 and -0.084 .

1.3.2 Phase-known vs. phase-unknown linkage data

Three-generation pedigrees with phase-known meioses are far more informative for linkage studies than the phase-unknown meioses observable in two-generation pedigrees. To illustrate this point, lod scores from a phase-unknown z_1 4 : 1 family are compared in Table 1.8 with phase-known families with either four recombinants and one non-recombinant, or one recombinant and four non-recombinants. For this reason, careful consideration of the families available for study at the outset of the project and selection of three-generation families likely to yield phase-known linkage data, may give great savings of time and effort in the long run.

1.3.3 Maximizing the informativeness of linkage studies

1.3.3.1 Choice of markers

The early human linkage studies had relatively few genetic traits available for investigation. These included several genetic diseases together with the blood group, red and white blood cell isozyme and serum markers. Using this limited supply of rather uninformative markers some remarkably sophisticated linkage maps were constructed for a few localized regions of chromosomes. With the advent of restriction fragment length polymorphisms (RFLP) [28] and the possibility of finding polymorphic markers for mapping in the intergenic regions previously devoid of markers, Solomon and Bodmer [29] and Botstein *et al.* [30] proposed that genetic maps spanning entire chromosomes could be constructed with varying levels of resolution depending on the number and informativeness of markers available. For this purpose Botstein *et al.* [30] introduced the concept of the polymorphism information content (PIC) value of a genetic marker in place of heterozygosity for predicting its informativeness in linkage studies. Whereas heterozygosity simply estimates the frequency of heterozygotes for a genetic marker in a population, the PIC value estimates the frequency of informative matings for that marker and takes into account the fact that half the progeny of matings of the type $A1A2 \times A1A2$ will also be heterozygous and therefore uninformative for linkage.

RFLPs are usually diallelic and therefore their PIC values are low, in the range of 0.2–0.4. In 1989, Weber

Number of siblings scored	NR : Rec	θ			
		0.01	0.05	0.10	0.30
1	1 : 0	+0.297	+0.279	+0.255	+0.146
	0 : 1	-1.699	-1.000	-0.699	-0.222
2	2 : 0	+0.593	+0.558	+0.510	+0.292
	1 : 1	-1.402	-0.721	-0.444	-0.076
	0 : 2	-3.398	-2.000	-1.398	-0.444
3	3 : 0	+0.890	+0.837	+0.765	+0.438
	2 : 1	-1.106	-0.442	-0.189	+0.070
	1 : 2	-3.101	-1.721	-1.143	-0.298
	0 : 3	-5.097	-3.000	-2.097	-0.666
4	4 : 0	+1.187	+1.116	+1.020	+0.584
	3 : 1	-0.809	-0.163	+0.066	+0.216
	2 : 2	-2.805	-1.442	-0.888	-0.152
	1 : 3	-4.800	-2.721	-1.842	-0.520
	0 : 4	-6.796	-4.000	-2.796	-0.888
5	5 : 0	+1.483	+1.395	+1.275	+0.730
	4 : 1	-0.512	+0.116	+0.321	+0.362
	3 : 2	-2.508	-1.163	-0.633	-0.006
	2 : 3	-4.504	-2.442	-1.587	-0.374
	1 : 4	-6.499	-3.721	-2.541	-0.742
	0 : 5	-8.495	-5.000	-3.495	-1.110
6	6 : 0	+1.780	+1.674	+1.530	+0.876
	5 : 1	-0.216	+0.395	+0.576	+0.508
	4 : 2	-2.211	-0.884	-0.378	+0.140
	3 : 3	-4.207	-2.163	-1.333	-0.228
	2 : 4	-6.203	-3.442	-2.286	-0.596
	1 : 5	-8.198	-4.721	-3.240	-0.964
	0 : 6	-10.194	-6.000	-4.194	-1.332
7	7 : 0	+2.077	+1.953	+1.785	+1.022
	6 : 1	+0.081	+0.674	+0.831	+0.654
	5 : 2	-1.915	-0.605	-0.123	+0.286
	4 : 3	-3.910	-1.884	-1.077	-0.082
	3 : 4	-5.893	-3.163	-2.031	-0.450
	2 : 5	-7.902	-4.442	-2.985	-0.818
	1 : 6	-9.897	-5.721	-3.939	-1.186
	0 : 7	-11.893	-7.000	-4.893	-1.554
8	8 : 0	+2.373	+2.230	+2.042	+1.169
	7 : 1	+0.378	+0.951	+1.088	+0.801
	6 : 2	-1.618	-0.327	+0.134	+0.433
	5 : 3	-3.614	-1.606	-0.821	+0.065
	4 : 4	-5.609	-2.885	-1.775	-0.303
	3 : 5	-7.605	-4.164	-2.729	-0.671
	2 : 6	-9.600	-5.442	-3.683	-1.039
	1 : 7	-11.596	-6.721	-4.638	-1.407
	0 : 8	-13.592	-8.000	-5.592	-1.775

Table 1.5 Equivalent lod scores for various values of the recombination fraction, θ , for phase-known parents.

NR, non-recombinant; Rec, recombinant.

Table 1.6 z_1 lod scores for various values of the recombination fraction, θ .

Number of siblings scored	z_1	θ			
		0.01	0.05	0.10	0.30
2	2:0	+0.292	+0.285	+0.215	+0.064
	1:1	-1.402	-0.721	-0.444	-0.076
3	3:0	+0.589	+0.535	+0.465	-0.170
	2:1	-1.402	-0.721	-0.444	-0.076
4	4:0	+0.886	+0.814	+0.720	+0.298
	3:1	-1.110	-0.464	-0.229	-0.011
	2:2	-2.805	-1.442	-0.887	-0.151
5	5:0	+1.182	+1.093	+0.975	+0.436
	4:1	-0.813	-0.186	+0.022	+0.095
	3:2	-2.805	-1.442	-0.887	-0.151
6	6:0	+1.479	+1.371	+1.231	+0.578
	5:1	-3.527	+0.093	+0.276	+0.222
	4:2	-2.512	-1.185	-0.673	-0.087
	3:3	-4.207	-2.164	-1.331	-0.227
7	7:0	+1.776	+1.650	+1.486	+0.723
	6:1	-0.220	+0.371	+0.532	+0.360
	5:2	-2.216	-0.907	-0.422	+0.019
	4:3	-4.207	-2.164	-1.331	-0.227
8	8:0	+2.072	+1.929	+1.741	+0.868
	7:1	+0.077	+0.650	+0.787	+0.503
	6:2	-1.919	-0.629	-0.167	+0.146
	5:3	-3.915	-1.906	-1.116	-0.163
	4:4	-5.609	-2.885	-1.775	-0.303

Table 1.7 z_2 lod scores for various values of the recombination fraction, θ .

Number of siblings scored	z_2	θ			
		0.01	0.05	0.10	0.30
2	2:0:0:0	+0.044	+0.037	+0.030	+0.008
	0:2:0:0	+0.292	+0.258	+0.215	+0.064
	1:1:0:0	-0.168	-0.137	-0.107	-0.024
	1:0:1:0	-0.049	-0.041	-0.032	-0.008
	1:0:0:1	+0.121	+0.104	+0.084	+0.023
	0:1:0:1	-1.402	-0.721	-0.444	-0.076
3	3:0:0:0	+0.121	+0.104	+0.084	+0.023
	0:3:0:0	+0.589	+0.535	+0.465	+0.170
	2:1:0:0	-0.331	-0.260	-0.191	-0.040
	2:0:1:0	-0.049	-0.041	-0.032	-0.008
	2:0:0:1	+0.242	+0.212	+0.175	+0.051
	1:2:0:0	+0.121	+0.104	+0.084	+0.023
	1:0:0:2	+0.415	+0.371	+0.315	+0.103
	1:1:1:0	-0.049	-0.041	-0.032	-0.008
	1:1:0:1	-1.402	-0.721	-0.444	-0.076
4	0:2:0:1	-1.402	-0.721	-0.444	-0.076
	4:0:0:0	+0.218	+0.190	+0.156	+0.044
	0:4:0:0	+0.885	+0.814	+0.720	+0.298
	3:1:0:0	-0.487	-0.361	-0.253	-0.049
	3:0:1:0	-0.005	-0.004	-0.002	0.000
	3:0:0:1	+0.364	+0.323	+0.271	+0.084

Table 1.7 Continued.

Number of siblings scored	z_2	θ			
		0.01	0.05	0.10	0.30
4	1:3:0:0	+0.417	+0.380	+0.331	+0.118
	0:3:0:1	-1.110	-0.464	-0.229	-0.011
	0:3:1:0	+0.712	+0.649	+0.568	+0.217
	2:2:0:0	-0.051	-0.049	-0.044	-0.014
	2:0:2:0	-0.098	-0.082	-0.064	-0.016
	2:0:0:2	+0.538	+0.485	+0.417	+0.144
	0:2:0:2	-2.753	-1.442	-0.887	-0.151
	2:1:1:0	-0.217	-0.178	-0.136	-0.032
	2:1:0:1	-1.358	-0.684	-0.414	-0.068
	2:0:1:1	+0.072	+0.063	+0.052	+0.015
	1:2:1:0	+0.243	+0.217	+0.183	+0.057
	1:2:0:1	-1.570	-0.858	-0.548	-0.100
	0:2:1:1	-1.282	-0.617	-0.360	-0.053
	1:1:1:1	-1.451	-0.762	-0.476	-0.084
5	5:0:0:0	+0.327	+0.288	+0.240	+0.072
	0:5:0:0	+1.182	+1.093	+0.975	+0.436
	4:1:0:0	-0.630	-0.431	-0.286	-0.051
	4:0:1:0	+0.072	+0.063	+0.052	+0.015
	4:0:0:1	+0.486	+0.435	+0.370	+0.122
	0:4:1:0	+1.008	+0.928	+0.823	+0.349
	0:4:0:1	-0.813	-0.186	+0.022	+0.095
	1:4:0:0	+0.714	+0.659	+0.585	+0.240
	3:2:0:0	-0.223	-0.201	-0.168	-0.046
	3:0:2:0	-0.098	-0.082	-0.064	-0.016
	3:0:0:2	+0.661	+0.598	+0.519	+0.189
	0:3:2:0	+0.834	+0.763	+0.670	+0.266
	0:3:0:2	-2.805	-1.442	-0.887	-0.151
	2:3:0:0	+0.245	+0.226	+0.197	+0.068
	3:1:1:0	-0.380	-0.301	-0.223	-0.048
	3:1:0:1	-1.282	-0.617	-0.360	-0.053
	3:0:1:1	+0.193	+0.171	+0.143	+0.043
	1:3:0:1	-1.282	-0.617	-0.360	-0.053
	0:3:1:1	-0.987	-0.350	-0.128	+0.027
	2:2:1:0	+0.072	+0.063	+0.052	+0.015
	2:2:0:1	-1.734	-0.981	-0.635	-0.116
	2:1:2:0	-0.098	-0.082	-0.064	-0.016
	2:0:1:2	+0.366	+0.330	+0.283	+0.095
	1:2:0:2	-2.805	-1.442	-0.887	-0.151
	2:1:1:1	-1.451	-0.762	-0.476	-0.084
	1:2:1:1	-1.451	-0.762	-0.476	-0.084
6	6:0:0:0	+0.442	+0.393	+0.331	+0.104
	0:6:0:0	+1.479	+1.371	+1.231	+0.578
	5:0:0:1	+0.610	+0.548	+0.471	+0.163
	5:0:1:0	+0.169	+0.149	+0.124	+0.036
	5:1:0:0	-0.749	-0.462	-0.287	-0.044
	1:5:0:0	+1.011	+0.938	+0.841	+0.376
	0:5:1:0	+1.305	+1.207	+1.078	+0.489
	0:5:0:1	-0.517	+0.093	+0.276	+0.222
	4:2:0:0	-0.394	-0.349	-0.284	-0.071
	4:0:2:0	-0.054	-0.044	-0.034	-0.008
	4:0:0:2	+0.783	+0.712	+0.621	+0.235
	0:4:2:0	+1.131	+1.042	+0.925	+0.401
	0:4:0:2	-2.512	-1.185	-0.673	-0.087
	2:4:0:0	+0.542	+0.504	+0.451	+0.184
	4:1:1:0	-0.536	-0.402	-0.285	-0.057
	4:1:0:1	-1.184	-0.531	-0.288	-0.032

Continued.

Table 1.7 Continued.

Number of siblings scored	z_2	θ			
		0.01	0.05	0.10	0.30
6	4:0:1:1	+0.315	+0.282	+0.239	+0.077
	0:4:1:1	-0.691	-0.072	+0.124	+0.141
	1:4:0:1	-0.985	-0.341	-0.113	+0.042
	1:4:1:0	+0.837	+0.773	+0.688	+0.290
	3:3:0:0	+0.074	+0.071	+0.064	+0.021
	3:0:3:0	-0.147	-0.123	-0.096	-0.023
	3:0:0:3	+0.957	+0.877	+0.773	+0.315
	0:3:0:3	-4.207	-2.164	-1.331	-0.227
	3:2:1:0	-0.100	-0.090	-0.076	-0.022
	3:2:0:1	-1.890	-1.082	-0.697	-0.125
	3:1:2:0	-0.266	-0.219	-0.168	-0.039
	3:1:0:2	-1.038	-0.398	-0.172	+0.009
	3:0:1:2	+0.489	+0.444	+0.385	+0.136
	3:0:2:1	+0.023	+0.022	+0.020	+0.007
	0:3:2:1	-0.864	-0.237	-0.027	+0.069
	0:3:1:2	-2.684	-1.339	-0.803	-0.129
	1:3:2:0	+0.663	+0.608	+0.536	+0.209
	1:3:0:2	-2.972	-1.579	-0.992	-0.175
	2:3:1:0	+0.368	+0.339	+0.299	+0.110
	2:3:0:1	-1.453	-0.770	-0.488	-0.090
	3:1:1:1	-1.407	-0.725	-0.446	-0.076
	1:3:1:1	-1.159	-0.505	-0.261	-0.019
	2:2:2:0	+0.194	+0.176	+0.151	+0.049
	2:2:0:2	-2.761	-1.405	-0.858	-0.144
	2:2:1:1	-1.619	-0.899	-0.580	-0.107
	2:1:2:1	-1.500	-0.803	-0.508	-0.091
	2:1:1:2	-1.331	-0.658	-0.392	-0.061
	1:2:1:2	-2.854	-1.483	-0.919	-0.159

Table 1.8 Comparison of two- and three-generation lod scores between families with five offspring.

Type of lod score		Recombination fraction			
		0.01	0.05	0.10	0.30
Two-generation data	$z_4:1$	-0.813	-0.186	+0.022	+0.095
Three-generation data	4Rec:1NR	-6.499	-3.721	-2.541	-0.742
	1Rec:4NR	-0.512	+0.116	+0.321	+0.362

NR, non-recombinant; Rec, recombinant.

and May [31] and Litt and Luty [32] introduced methods for typing a new class of highly polymorphic markers called microsatellites or short tandem repeats (STRs) (see Chapter 5). These STR markers may have so many alleles that almost every individual tested is informative for linkage, and therefore their PIC values are typically in the range 0.7–0.9.

STR markers are very common in the human genome at an estimated density of around one every 10–100 kb of genomic DNA. This by far exceeds the resolution of linkage analysis, which is limited by

the availability of informative pedigrees in which recombination can be detected. This is generally considered to be about 0.1cM or approximately 100 kb of genomic DNA if the rule of thumb conversion of 1 cM corresponding to 1 Mb of DNA is applied.

1.3.3.2 Haplotype analysis

The existence of localized high-density linkage maps permits the construction of haplotypes. A haplotype is a combination of alleles at a number of very closely linked loci which usually segregate

together through a pedigree. As the genetic length of the haplotype increases, then the chance of recombination occurring to disturb the haplotype also increases. Haplotypes can be constructed in pedigrees using data from two or more close markers in the combination which permits the least number of cross-overs. Alternatively, where a large number of markers are being used or the pedigrees are large, it may be preferable to use linkage analysis programs such as LINKAGE to assist haplotype construction.

Haplotype data are important in linkage studies because they increase the amount of available information by, in effect, increasing the number of alleles. In addition, where a new marker—for example, a disease locus, is being added to an established linkage map—recombination events within the haplotype can be used to localize the marker on one side or the other of each recombination event. Methods for handling multiple point mapping data are discussed in more detail in Chapter 3.

1.4 Sex linkage

So far linkage has been considered for loci on the autosomes—that is, not on the sex chromosomes. The X chromosome is known to carry very many genes of clinical and scientific importance and so linkage mapping on this chromosome is of great relevance.

X-linked traits show characteristic patterns of transmission through pedigrees. X-linked dominant traits such as hypophosphataemic rickets are observed in males and females. All the daughters of affected males but none of their sons exhibit the trait. Affected mothers transmit X-linked dominant traits to half their sons and daughters equally. With X-linked recessive traits such as haemophilia A, unaffected carrier mothers transmit the trait to half their sons, and half their daughters are unaffected carriers. Affected fathers have only unaffected sons, but all their daughters are unaffected carriers of the trait. In the rare situation of homozygous affected mothers, all their sons are affected and all their daughters are carriers.

As a result of these characteristic inheritance patterns with a sex bias amongst the affected offspring, simple inspection of pedigrees should confirm whether or not a trait is X linked. Because the prior odds on two sex-linked loci being linked are so much higher than when dealing with autosomal loci, the statistical evidence needed to confirm linkage is less stringent. Therefore a lod score of +2 between two X-linked loci is usually considered adequate to demonstrate their genetic

linkage, whereas for autosomal loci a lod score of +3 is generally required.

Identification of genes involved in malignant hyperthermia susceptibility

Malignant hyperthermia (MH) is a potentially lethal reaction to inhalation of anaesthetics and is one of the most common causes of death in otherwise healthy individuals undergoing general anaesthesia. An autosomal dominant mode of inheritance for susceptibility to MH was proposed by Denborough *et al.* [34]. They also suggested that the MH susceptibility (MHS) phenotype shows incomplete penetrance, because one individual in this pedigree had transmitted the susceptibility gene to her offspring although she herself had not experienced an MH crisis during general anaesthesia. The frequency of MHS has been estimated at one in 5000 in the UK population [35], although the frequency of MH crises is considerably lower because of the reduced penetrance. Presymptomatic diagnosis of MHS is possible using an *in vitro* contracture test (IVCT) [36,37]. This test involves measuring the strength of contracture of living muscle fibre bundles exposed to halothane or caffeine and therefore is highly invasive and expensive.

Genetic investigations of MHS were stimulated by the mapping of a gene in pigs (*Hal*) that causes a condition similar to MH [38,39]. The *Hal* gene was localized to a region of the porcine gene map that shows conservation across a remarkably wide range of species, including humans [40]. In humans, the syntenic region lies on the long arm of chromosome 19q12-q13.1. Evidence in humans for linkage to markers in this region [41,42] identified a plausible candidate gene, *RYR1*, for the calcium-sensitive calcium release channel, or ryanodine receptor, of skeletal muscle sarcoplasmic reticulum. A causative mutation, C1843T, resulting in a cysteine for arginine substitution at residue 615, was identified in this gene in halothane-sensitive pigs in all breeds affected [43]. The equivalent mutation, C1840T, was subsequently identified in a few human families [44] and may account for around 5% of human MH susceptibility [35].

Further linkage studies in MHS families revealed that possibly as much as 50% of MH is not linked to *RYR1*, indicating that high levels of genetic heterogeneity exist in this condition, although it is impossible to distinguish different phenotypic groups between the *RYR1*-linked and unlinked families. Genome searching has localized two further MHS loci, *MHS3* on chromosome 7q [45] and *MHS4* on chromosome 3q [46]. An apparent MHS locus on chromosome 17q is likely to be due to misdiagnosis of hyperkalaemic periodic paralysis caused by mutations at the *SCN4A* locus located in this region [47]. Linkage studies have only identified *MHS3* and *MHS4* as the defective genes in single MHS families. Therefore the question arises as to how many further MHS genes are still to be identified, and will they be 'private' gene defects occurring only in single isolated families?

Case Study 1.1

Because males are hemizygous, having only one copy of X-linked alleles, the phase of fathers is known if they have been typed for the markers being considered. Furthermore, it is common to be able to establish phase for the heterozygous mother because in the grandparental generation only three alleles are present, two from grandmother and one from grandfather, compared with four alleles when dealing with autosomal loci.

Except for loci located within the pseudoautosomal regions (*PAR1* and *PAR2*) at the tips of each

arm of the X chromosome, which pair with their counterparts on the Y chromosome in meiosis, genetic recombination on the X chromosome is only observed in females. This simplifies establishing haplotypes for linkage because the possibilities for crossing-over are reduced.

For these reasons, studies on the X chromosome provided the first example of a human linkage map spanning the entire length of a chromosome. It remains the most highly resolved of all the human linkage maps.

.....

Troubleshooting

Phenotype of interest does not segregate in a Mendelian pattern

This may mean one of the following:

- *The trait is not genetically determined, i.e. it is not determined by Mendelian genes but it may be determined by mitochondrial genes.*
- *The trait is not due to a single gene (see Chapter 2).*
- *The trait shows variable expressivity. If you suspect this, re-evaluate phenotypes of family members. For example, are there very mildly affected individuals who have been counted as unaffected? Are there individuals who have died before their disease status could be determined?*
- *The trait shows variable age of onset and/or reduced penetrance. Use 'age of onset' curves for estimating risk of being affected in unaffected family members.*
- *Phenocopies occur in the population. Is a test possible to discriminate between gene-caused phenotype and the non-genetic phenotype? Estimate the frequency of the phenotype in the population in non-familial cases and use this to weight the risk of being a gene carrier for affected individuals.*
- *De novo mutation events commonly give rise to phenotype. Reinterpret pedigrees allowing for new mutation events.*

Family structure is not as recorded (i.e. non-paternity, adoption)

- *Check family structure by inspecting microsatellite marker linkage data. If pedigree is not as supposed, redraw pedigree leaving out the non-fit individuals from analysis if necessary.*

Linkage markers do not segregate in a Mendelian pattern

- *Samples may have been mixed up. Discard samples and collect fresh blood samples from family.*
- *Family structure is not as recorded (i.e. non-paternity, adoption).*

Redraw pedigree leaving out the non-fit individuals from analysis if necessary.

- *Alleles have been misinterpreted. Reinspect and reinterpret raw data.*

Linkage markers segregate with disease in some families but not in others

Defects in different genes underlie the phenotype (i.e. genetic heterogeneity).

- *Collect marker data from large families only. Do not pool linkage data because the unlinked kinships will negate any possible positive data. Use heterogeneity tests such as HOMOG [33].*

Haplotypes cannot be constructed

- *Loci are not linked.*
- *Order of loci is incorrect.*

Distances between loci are incorrect

- *For all these, check genetic map for markers under investigation.*

.....

References

1 Sutton, W.S. (1903) The chromosomes in heredity. *Biol Bull.* **4**, 231–251.

2 deVries, H. (1903) *Befruchtung und Bastardiening* (Leipzig). [English translation in *Intracellular pangene-sis including a paper on fertilization and hybridization* (Open Court Publ. Co., Chicago, 1910), 217–263.]

3 Bateson, W., Saunders, E.R. & Punnett, R.C. (1905) Experimental studies in the physiology of heredity. *Rpt. Evol. Communication R. Soc. Lond.* **2**, 1–55, 80–99.

4 Morgan, T.H. (1911) An attempt to analyse the constitution of the chromosomes on the basis of sex-limited inheritance in *Drosophila*. *J. Exp. Zool.* **11**, 365–414.

5 Morgan, T.H. & Cattell, E. (1912) Data for the study of sex-linked inheritance in *Drosophila*. *J. Exp. Zool.* **13**, 79–101.

6 Sturtevant, A.H. (1913) The linear arrangement of six sex-linked factors in *Drosophila*, as shown by the mode of association. *J. Exp. Zool.* **14**, 43–59.

7 Janssens, F.A. (1909) Spermatogenesis dans le bactra-ciens V. La theorie de la chiasmotypie. Nouvelles inter-pretations des cineses de maturation. *Cellule* **25**, 387–411.

8 Creighton, H.S. & McClintock, B. (1931) A correlation of cytological and genetical crossing-over in *Zea mays*. *Proc. Natl Acad. Sci. USA* **17**, 492–497.

9 Muller, H.J. (1916) The mechanism of crossing-over. *Am. Nat.* **50**, 191–221, 284–305, 350–366, 421–434.

10 Haldane, J.B.S. (1919) The combination of linkage values and the calculation of distances between the loci of linked factors. *J. Genet.* **8**, 229–309.

11 Kosambi, D.D. (1944) Estimation of map distances from recombination values. *Annls Eugen.* **12**, 172–175.

12 Carter, T.C. & Falconer, D.S. (1951) Stocks for detecting linkage in the mouse and the theory of their design. *J. Genet.* **50**, 307–323.

13 Sturt, E. (1976) A mapping function for human chromosomes. *Annls Hum. Genet.* **40**, 147–163.

14 McDermott, A. (1973) The frequency and distribution of chiasmata in man. *Annls Hum. Genet.* **37**, 13–20.

15 Rao, D.C., Morton, N.E., Lindsten, J. *et al.* (1977) A mapping function for man. *Hum. Hered.* **27**, 99–104.

16 Hultèn, M. (1974) Chiasma distribution at diakinesis in the normal human male. *Hereditas* **76**, 55–78.

17 Vasen, H.F.A., Mecklin, J.-P., Meera Khan, P. & Lynch, H.T. (1991) The international collaboration group on hereditary non-polyposis colon cancer (ICG-HNPCC). *Dis. Colon Rectum* **34**, 424–425.

18 Haldane. (1935) The rate of spontaneous mutation of a human gene. *J. Genet.* **31**, 317–326.

19 Morton, N.E. (1955) Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* **7**, 277–318.

20 Maynard-Smith, S., Penrose, I.S. & Smith, C.A.B. (1961)

Mathematical Tables for Research Workers in Human Genetics (J.A. Churchill, London).

21 Ott, J. (1974) Estimation of the recombination in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am. J. Hum. Genet.* **26**, 588–597.

22 Lathrop, G.M. & Lalouel, J.M. (1984) Easy calculations of lod scores and genetic risks on small computers. *Am. J. Hum. Genet.* **36**, 460–465.

23 Smith, C.A.B. & Sturt, E. (1976) The peak of the likelihood curve in linkage testing. *Ann. Hum. Genet.* **39**, 423–426.

24 Cook, P.J.L. (1965) The Lutheran-secretor recombination fraction in man: a possible sex difference. *Ann. Hum. Genet.* **28**, 393–401.

25 Renwick, J.H. & Schultze, J. (1965) Male and female recombination fractions for the nail-patella: ABO linkage in man. *Ann. Hum. Genet.* **28**, 379–392.

26 Attwood, J. & Bryant, S. (1988) A computer program to make analysis with LIPED and LINKAGE easier to perform and less prone to input errors. *Ann. Hum. Genet.* **52**, 259.

27 Race, R.R. & Sanger, R. (1975) *Blood Groups in Man* (6th edn, Blackwell Scientific Publications, Oxford).

28 Kan, Y.W. & Dozy, A.M. (1978) Polymorphism of DNA sequence adjacent to human β -globin structural gene; relationship to sickle mutation. *Proc. Natl Acad. Sci. USA* **75**, 5631–5635.

29 Solomon, E. & Bodmer, W.F. (1979) Evolution of sickle-cell variant gene. *The Lancet*, April **28**, 923.

30 Botstein, D., White, R.L., Skolnick, M. & Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314–331.

31 Weber, J.L. & May, P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388–96.

32 Litt, M. & Luty, J.A. (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44**, 397–401.

33 Ott, J. (1991) *Analysis of Human Genetic Linkage* (Johns Hopkins Press, Baltimore).

34 Denborough, M.A., Forster, F.J.A., Lovell, J.J.H. *et al.* (1962) Anaesthetic deaths in a family. *Brit. J. Anaesth.* **34**, 395–396.

35 West, S.P. (1996) The genetics of malignant hyperthermia. In *Hyperthermia and Hypermetabolic Disorders* (Ellis, F.R. & Hopkins, P.M., eds) 184–201 (Cambridge University Press, Cambridge).

36 European Malignant Hyperthermia Group (1984) A protocol for the investigation of malignant hyperpyrexia (MH) susceptibility. *Brit. J. Anaesth.* **56**, 1267–1269.

37 European Malignant Hyperthermia Group (1985) Laboratory diagnosis of malignant hyperthermia susceptibility. *Brit. J. Anaesth.* **57**, 1038.

38 Archibald, A.L. & Imlah, P. (1985) The halothane sensitivity locus and its linkage relationships. *Anim. Blood Groups Biochem. Genet.* **16**, 253–263.

39 Harbitz, I., Chowdhary, B., Thomsen, P.D. *et al.* (1990) Assignment of the porcine calcium release channel gene, a candidate for the malignant hyperthermia locus, to the 6p11-q21 segment of chromosome 6. *Genomics* **8**, 243–248.

40 Eiberg, H., Nielsen, L.S., Gahna, B. *et al.* (1989) Exclusion data for the α -B glycoprotein (GA1B) polymorphism. *Cytogenet. Cell Genet.* **51**, 994.

41 McCarthy, T.V., Healy, J.M.S., Heffron, J.J.A. *et al.* (1990) Localization of the malignant hyperthermia susceptibility locus to chromosome 19q12-q13.2. *Nature* **343**, 562–564.

42 MacLennan, D.H., Duff, C., Zoraa, F. *et al.* (1990) Ryanodine receptor is a candidate for predisposition to malignant hyperthermia. *Nature* **343**, 559–561.

43 Fujii, J., Otsu, K. & Zorzato, F. *et al.* (1991) Identification of a mutation in the porcine ryanodine receptor that is associated with malignant hyperthermia. *Science* **253**, 448–451.

44 Gillard, E.F., Otsu, K., Fujii, J. *et al.* (1991) A substitution of cysteine for arginine 614 in the ryanodine receptor is potentially causative of human malignant hyperthermia. *Genomics* **11**, 751–755.

45 Iles, D.E., Lehmann-Horn, F., Scherer, S.W. *et al.* (1994) Localization of the gene encoding the α 2/ δ -subunits of the L-type voltage-dependent calcium channel to chromosome 7q and analysis of the segregation of flanking markers in malignant hyperthermia susceptible families. *Human Molec. Genet.* **3**, 969–975.

46 Sudbrak, R., Procaccio, V., Klautsitzer, M. *et al.* (1995) Mapping of a further malignant hyperthermia susceptibility locus to chromosome 3q13.1. *Am. J. Hum. Genet.* **56**, 684–691.

47 Iazzo, P.A. & Lehmann-Horn, F. (1995) Anesthetic complications in muscle disorders. *Anesthesiology* **82**, 1093–1096.

Chapter 2

Genetic analysis of complex traits

D. Timothy Bishop

ICRF Genetic Epidemiology Laboratory, St James’s University Hospital, Leeds LS9 7TF, UK

2.1	Introduction, 28	2.5	Sampling problems, 37
2.2	Mendelian trait with covariates, 28	2.6	Issues of analysis, 38
2.3	Genetic heterogeneity, 33	2.6.1	Looped pedigrees, 38
2.3.1	Phenotypic heterogeneity, 33	2.6.2	Dealing with large numbers of marker alleles, 39
2.3.2	Analysis of linkage heterogeneity, 34	2.7	General discussion, 40
2.4	Complex traits with no clear mode of inheritance, 35		References, 40

2.1 Introduction

Linkage mapping has proved impressively successful in mapping genes determining Mendelian diseases. The genes for Huntington’s disease [1], adenomatous polyposis coli [2,3], cystic fibrosis [4], and many other diseases were mapped in this way, by repeatedly looking for cosegregation of DNA markers with the disease. The basic approach to linkage mapping and the historical perspective is outlined in Chapter 1, while the development of genetic maps covering the whole genome is discussed in Chapter 3. The success in identifying the genes for these ‘rare’ diseases has vindicated the positional cloning approach and has led to an interest in the more common diseases that are not Mendelian in segregation. The impetus for such studies comes from both a scientific and a public health perspective since the majority of diseases show family aggregation, suggesting a genetic contribution to their aetiology. For instance, cardiovascular diseases, cancer and psychiatric diseases occur more frequently in the relatives of patients than in the general population. There are, of course, other non-genetic explanations for such familial aggregations, but often the increased risk of disease in the relatives is so great that known non-genetic factors cannot account for it [5].

In this chapter, the term ‘complex’ will be used to describe diseases that are not inherited as simple Mendelian traits. The diseases may be complex in many different ways, and so this chapter can by no means be exhaustive. Approaches to dealing with some of these complexities will be discussed. However, it is impossible to be prescriptive in these circumstances, and studies must rely on the state of knowledge for the particular disease; it is hoped that the following may give some guidance when planning studies.

This chapter focuses initially on those diseases that do have a Mendelian component, while diseases without any such clues to their aetiology are discussed later.

2.2 Mendelian trait with covariates

A natural place to start the discussion is with diseases that are clearly due to one or more genes, each of which is sufficient to determine susceptibility. For instance, many diseases have a Mendelian component, but other factors may mask the inheritance in individual families or only a subset of families of cases may show evidence of a predisposition. The Mendelian component is recognized by the occasional identification of a family

with both close and distant relatives affected, and in which the ‘pattern’ of relationship among the affected individuals is consistent with the inheritance of a single gene. Such families will be more notable if the inherited trait is due to a dominant gene.

There are often only limited numbers of affected relatives in these families because the disease is predominantly expressed in a subgroup of the population, such as a particular age range or gender. For instance, a disease that is expressed (or primarily expressed) in one gender only, but which is dominantly inherited or in which onset does not occur in childhood, will appear like this. For this discussion, the examples will be taken from studies of breast cancer. In some families, susceptibility to breast cancer is inherited as an autosomal dominant. The onset is earlier than in the general population (often occurring when a woman is in her thirties or forties). In 1990, a gene for hereditary breast cancer was mapped to chromosome 17 by Hall *et al.* [6]; this gene is now called *BRCA1*. A collaborative study of families with a number of cases of early onset breast cancer were collected and published by Easton *et al.* [7]. Figure 2.1 shows four families from that analysis, slightly modified for this discussion. The pedigrees show the anticipated ‘dominant’ features of disease in most generations; the disease-associated mutations can be traced through fathers and unaffected mothers, but with a high risk to (female) carriers of the mutation.

In Chapter 1, the concept that there might not be a simple relationship between genotype and phenotype was introduced, but in the cases discussed in that chapter, each genotype determined the phenotype precisely (i.e. for a dominant disease or a recessive disease). We now introduce the more general concept of penetrance; penetrance is the probability that an individual is affected given their genotype at the disease-causing locus. For a single locus with three genotypes (alleles *A*, *a*), there are therefore three penetrance probabilities to be defined:

$$\begin{aligned} &P[\text{of being affected} \mid \text{person's genotype is } AA] \\ &P[\text{of being affected} \mid \text{person's genotype is } Aa] \\ &P[\text{of being affected} \mid \text{person's genotype is } aa] \end{aligned}$$

where $P[\dots]$ represents the probability of the event included in the parentheses, and the vertical bar indicates that the probability of being affected is when the person’s genotype is as specified. For example, for a dominant disorder without phenocopies (i.e. when an affected person has to have at least one copy of *A*):

$$\begin{aligned} P[\text{being affected} \mid AA] &= 1.0 \\ P[\text{being affected} \mid Aa] &= 1.0 \\ P[\text{being affected} \mid aa] &= 0.0 \end{aligned}$$

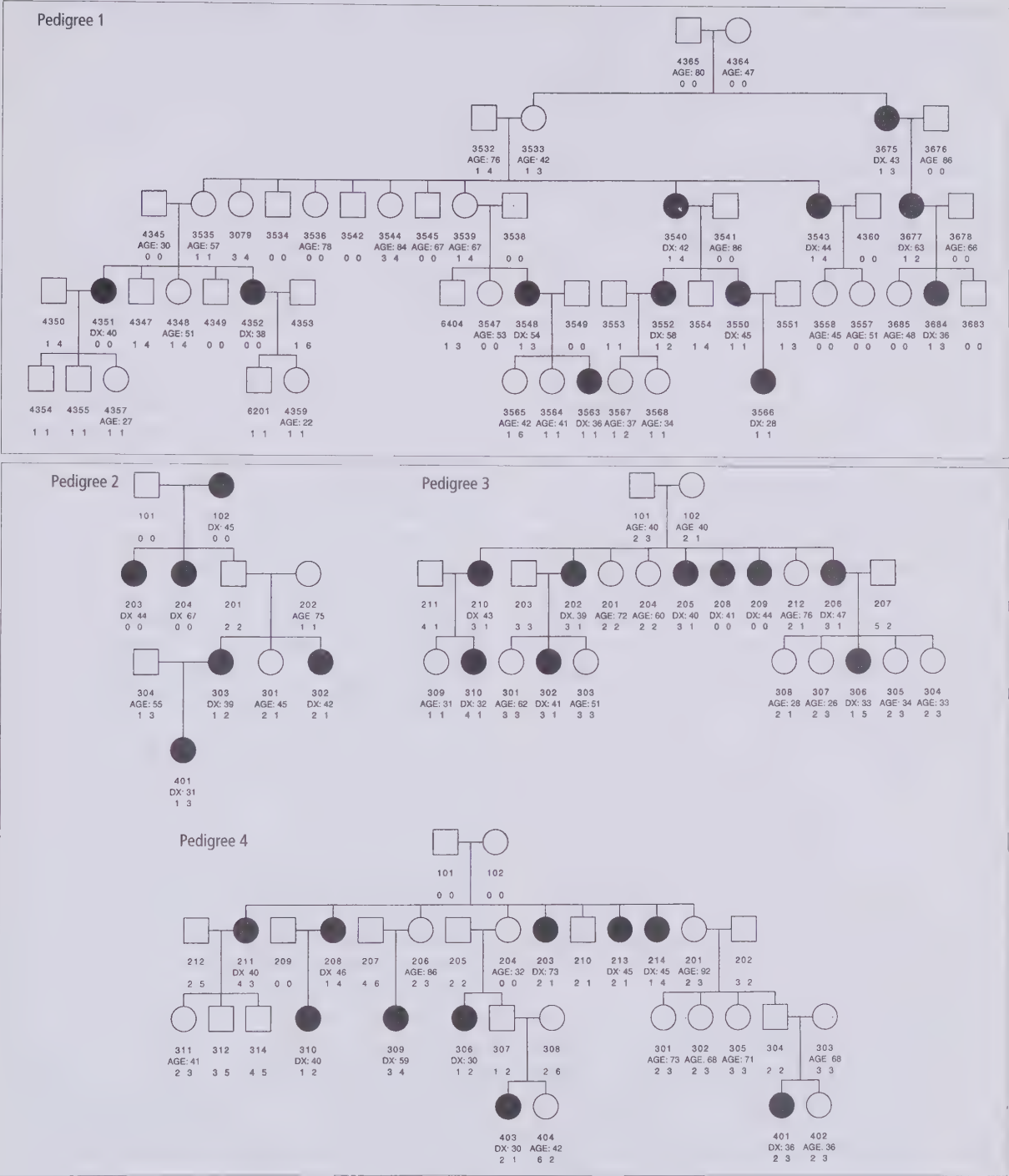


Fig. 2.1 Four families ascertained with breast cancer as part of the Breast Cancer Linkage Consortium [7]. Individuals shaded are affected with breast cancer, ages under the individuals show either current age if unaffected or age at presentation of disease (DX:) if

affected. Underneath each pedigree symbol is the person's identification number, their age (current or age at diagnosis) and their marker typing for a chromosome 17q marker adjacent to *BRCA1*.

For ease of notation, we leave ‘person’s genotype is’ from the formulae. This pattern applies to the disease adenomatous polyposis coli, for instance, if we concentrate on people aged over 15 years. Often, however, the probability of being affected depends on other factors; these are often sex and/or age but are not exclusively so. For instance; the probability of developing melanoma may depend on the level of exposure to the sun as well as on genotype.

For the breast cancer pedigrees shown in Fig. 2.1, assume that men *never* develop breast cancer (not entirely true but acceptable for this discussion), and that susceptibility to breast cancer is determined at least in part by a rare dominant gene. We consider the following penetrance probabilities as an example:

	AA	Aa	aa
Females	0.40	0.40	0.03
Males	0.00	0.00	0.00

This table is read in the following way: females with at least one copy of the *A* allele have a risk of 0.40 of developing breast cancer while other women have a risk of 0.03, males will not develop breast cancer whatever their genotype. In the jargon of the LINKAGE computer program (see Chapter 3) the males and females represent distinct ‘liability classes’ (i.e. groups with a different liability to disease). Each liability class represents a different level of exposure to the risk factor (i.e. gender in this case). This table of probabilities indicates that whatever the genotype, a male will not develop breast cancer (our assumption above) but that a female carrying either one or two copies of *A* has a 40% risk of developing breast cancer as compared to 3% if she does not have a copy of *A*; as we know that dominantly inherited susceptibility to breast cancer is rare, the risk for *aa* is essentially the general population risk of breast cancer.

We should note that the relevant risks for this analysis are not the lifetime risks of breast cancer, since these would only be applicable if all gene carriers in the pedigrees had lived through all of their years at risk. In fact, the penetrance probabilities should reflect the average risk of women of the ages depicted in Fig. 2.1 (which include both younger and older women).

In the above example, two liability classes are considered because it is clear that men and women have different risks of the phenotype (breast cancer in this example). In general, however, the risk of disease usually depends upon other factors than age and genotype. The most common such complicating factor for any analysis is age, as indicated by the difficulty of calculating the appropriate penetrance probabilities in the above example.

Before we consider such issues, there are two rules about defining liability classes for linkage analysis. These are:

- 1 each person in the pedigree must be classifiable into exactly one liability class;
- 2 within each liability class, all other known risk factors should have little effect on overall risk of disease.

Rule 1 must be followed (and is clearly followed in this instance, where gender determines liability), while failure of rule 2 will mean that linkage will be harder to detect in that the power of the study has been reduced.

For the breast cancer example from above, a more realistic assumption would be that the risk of breast cancer in both gene carriers and non-carriers is age-specific. The major problem is then the question of assigning probabilities to each genotype for each liability class. The optimal way to estimate such probabilities is to refer to some external source of information such as the results from segregation analysis.

For breast cancer there is a ready solution, which is to refer to the published segregation analyses of breast cancer, e.g. [8]. In this analysis, a single high-risk dominant gene was the best fitting single-gene model. The risks for gene carriers and non-carriers are given in Table 2.1. In the liability model, ages are considered as discrete rather than continuous factors. The figures in this table show the cumulative risk of developing breast cancer by genotype as a function of age.

One of the obvious practical issues at the beginning of any analysis is the need to define the numbers and features of the liability classes. For the breast cancer example described above, age might be considered as broken into decades, 20-year periods or even longer periods. It was convenient to use the published results from segregation analysis which used 10-year age periods. The important feature is that the liability classes show the important distinctions between the risk by genotype. There is no simple (or completely correct) answer to the question of the number of liability classes. In fact, the appropriate number for any analysis will depend upon the state of knowledge regarding the inheritance of the disease. Knowledge that provides informed estimates of the age (or other cofactor) effect on risk, if correct, can only enhance the ability to detect linkage.

Linkage analysis relies upon the specification of the mode of inheritance for calculations as to the evidence for linkage in specific families. Any information regarding the mode of inheritance can be included in the specification of the linkage model,

Table 2.1 The estimated cumulative probability of a female being affected with breast cancer by a given liability class.

Liability class	Age (years)	Cumulative probability for each genotype	
		AA/Aa	aa
1	20–29	0.02	0.0002
2	30–39	0.14	0.0027
3	40–49	0.38	0.0138
4	50–59	0.55	0.0275
5	60–69	0.67	0.0497
6	70–79	0.95	0.0798
7	80+	1.00	0.1254

The estimated frequency of *A* is 0.0033 [8].

and indeed should be included in the specification of the model. Often, there is some information about the mode of inheritance but not all of the details are apparent. For instance, it may well be clear that there is an autosomal dominant component to inherited susceptibility but the age-specific risks may not be known for gene carriers (or, indeed, sometimes for non-carriers) and the gene frequency may be a matter of considerable speculation. In these circumstances, it is usual to consider the dominant component and to estimate the appropriate risks for carriers in some reasonably systematic way (such as

the proportion of probable gene carriers affected with the disease). The issue to remember with respect to these ‘guesses’ as to mode of inheritance is that they should be regarded as preliminary only; as soon as linkage is identified, it should be possible to update the information on the estimates of the age-specific risks. For this reason, and because of the statistical problem of multiple testing, the liability classes and the associated penetrances should be decided upon prior to the statistical linkage analysis. The issue of multiple testing is discussed in Section 2.4.

Table 2.2 shows the linkage results by family and

Table 2.2 The linkage analysis results for the breast cancer families shown in Fig. 2.1 depending on the number of liability classes assumed in the analysis.

<i>Two liability classes</i>							
Putative recombination fraction between disease locus and marker locus							
	0.0001	0.01	0.05	0.1	0.2	0.3	0.4
PED 1	2.41	2.37	2.19	1.96	1.47	0.94	0.41
PED 2	–0.63	–0.61	–0.51	–0.41	–0.27	–0.16	–0.07
PED 3	1.57	1.54	1.41	1.23	0.87	0.49	0.15
PED 4	0.79	0.79	0.78	0.73	0.54	0.30	0.10
Total	4.14	4.10	3.87	3.52	2.61	1.57	0.59
<i>Eight liability classes</i>							
Putative recombination fraction between disease locus and marker locus							
	0.0001	0.01	0.05	0.1	0.2	0.3	0.4
PED 1	2.89	2.84	2.64	2.38	1.81	1.19	0.53
PED 2	–1.33	–1.19	–0.85	–0.63	–0.37	–0.21	–0.09
PED 3	1.29	1.31	1.32	1.25	0.98	0.60	0.22
PED 4	1.63	1.64	1.63	1.54	1.20	0.75	0.27
Total	4.48	4.60	4.74	4.54	3.62	2.34	0.98

Note that the results are more informative in the eight liability class analysis and that there is better distinction between the families in terms of which are linked and which are unlinked. PED 1–PED 4 are four pedigrees published as part of the Breast Cancer Linkage Consortium [7].

total when either two liability classes (one male, one female as described above), or eight liability classes (seven age-dependent classes for females, one for males), are assumed. As can be seen, the overall results are similar, but the effect is that three out of the four families are more informative (i.e. have lod scores that are farther away from 0.0) under the eight liability model than the two liability model. Even the simpler assumption (two liability classes) shows strong evidence about linkage, however. This is the general rule concerning liability classes — that is, the more appropriate modelling of penetrance leads to more informative linkage analyses.

One further issue should be mentioned in considering liability classes when dealing with age as a factor. This relates to the attempt to distinguish further between carriers and non-carriers. Consider once more the breast cancer example. For instance, suppose a woman at 50% prior risk of carrying a copy of *A* (i.e. one parent is known to carry *A*) is affected at age 85, then an application of Bayes theorem using the penetrance figures of Table 2.1 shows that the probability that she carries that copy of *A* is:

$$\frac{1.00 \times 0.5}{(1.00 \times 0.5) + (0.1254 \times 0.5)} = 0.89 \quad (2.1)$$

that is, she is very likely to carry the *A* allele. However, a simple examination of Table 2.1 shows that most women with the *A* allele have developed their breast cancer prior to that age; also, for the 80+ age group the probability that a gene carrier develops breast cancer is similar to the risk that a non-carrier develops breast cancer ($1.00 - 0.95 = 0.05$ for carriers vs. $0.1254 - 0.0798 = 0.046$ for non-carriers). The strength of the evidence from the linkage analysis that she carries *A* and is therefore informative for the segregation of the disease gene through the family is inflated; in other words, sporadic cases are likely to be classified as carriers, and if they show evidence of recombination which will happen 50% of the time simply by chance then they will falsely negate the evidence for linkage. The solution to this is to reconsider the definition of the liability classes in the linkage analysis. In the current linkage analysis programs, 'affected by the time the person has reached age group *x*' and 'unaffected and in age group *x*' are regarded as being the only two options for an individual in age group *x*. However, this ignores the information that an affected individual has developed the disease while in age group *x* rather than prior to that age group. For these analyses it would be more appropriate to consider:

- those unaffected *up to* age group *x*; and
- those affected *at* age group *x*.

Such a change makes linkage analysis analogous to the statistical technique of 'survival analysis' [9]. The two possible observations are not now complementary. This can be accommodated in LINKAGE by considering two sets of liability classes, one which relates to affected individuals and one which relates to unaffected individuals [10]. For unaffected individuals, the liability classes and associated probabilities are as described previously; for affected individuals, the risks associated with each age group are simply the risk of developing the disease in that age group. For instance, from Table 2.1, the risk of developing breast cancer in her sixties is 0.08 for a gene carrier ($0.67 - 0.55 = 0.08$) and 0.0222 ($0.0497 - 0.0275$) for a non-carrier. This modification will usually allow clearer definition of the carrier status of affected individuals, making the linkage analysis more informative overall [11].

One of the advantages of linkage analysis of diseases with a clear Mendelian component is that the analysis incorporates information on unaffected individuals while most other methods ignore such information (see above). There is of course information to be gained, especially if unaffected siblings carry discordant marker alleles to their affected siblings. Analysis of affected individuals only does not allow such a comparison and so can be less informative. The precise level of informativeness depends upon the actual risk that gene carriers develop the disease. To indicate the importance of this issue, consider a rare dominant disease (high-risk allele is labelled *D*, wild-type allele is labelled *d*) with an associated risk of cancer of (*t*) and suppose that the phase is known in the carrier grandparent (and hence parent in this example) as would be the case if the pedigree shown in Fig. 2.2 was a recently discovered branch of an extended family. There are four types of observations for the children:

- 1 affected having inherited the marker allele from the mutation-bearing chromosome of the parent ('high-risk allele', *D*);
- 2 affected having inherited the low-risk marker allele from the parent;
- 3 unaffected having inherited the low-risk marker allele from the carrier parent (these are consistent with linkage ('low-risk allele', *d*));
- 4 unaffected having inherited the high-risk allele from the parent.

Figure 2.2 shows the contribution of each child to the total linkage analysis (in this simple example, offspring contribute independently to the total lod score) supposing that the disease gene is at a recombination distance of 0.05 from the marker being considered here.

There are a number of simple observations

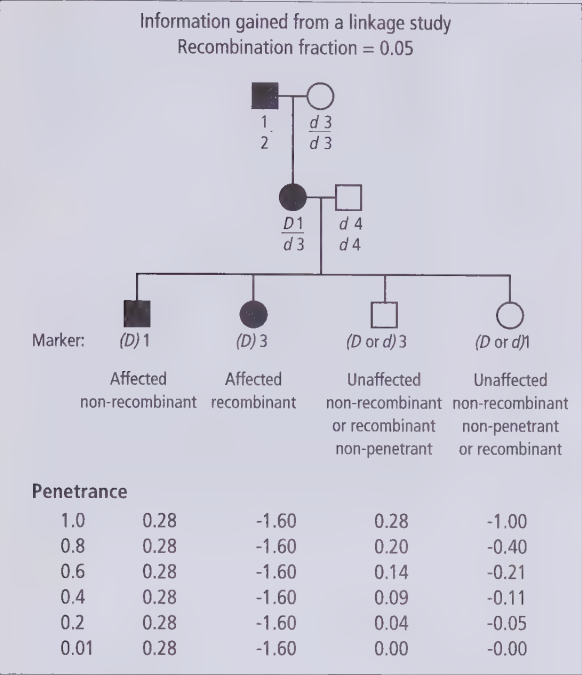


Fig. 2.2 In families segregating an inherited susceptibility, different observations have varied effects on the lod score. In this example, consider a rare dominant disease with partial penetrance in those carrying a susceptibility allele (*D*; *d* is the wild-type allele). The family structure shows that the mother carries the susceptibility and the marker allele in phase with *D* (allele 1 in this example). The four possible observations for the children are then affected or unaffected and carrier of high-risk marker allele (marker allele 1) or the low-risk allele (marker allele 3) from the segregating parent. In the children, the only information is the disease status and the marker allele, and *not* the most useful information which is whether or not they have inherited *D* from the mother.

(Fig. 2.2). First is that affected individuals are more informative than unaffected, except in the extreme case of a penetrance of $t=1.0$ in carriers. In this instance, unaffected and affected individuals are equally informative, since disease status is a perfect indicator of their underlying genotype. Second, the affected individuals contribute to the linkage analysis the same lod score independently of the value of t ; this is true whenever the risk of the disease is zero or close to zero in non-carriers and so is a feature of most analyses. The result would be less clear if the carrier status of the father were less sure than here. Third, the affected non-recombinant gives more information in favour of linkage than the affected recombinant gives against linkage. This is because recombination events occur in only one in 20 meioses in a linked family (on the basis of the assumption in Fig. 2.2 of a recombination fraction of 0.05) and so finding a clear recombinant individual

is evidence that this may not be a linked family. Finally, unaffected individuals give minimal evidence for or against linkage when the penetrance is low. More precisely, the uninformative individuals only give useful information when the penetrance is 0.8 or more which is not the case for many of the diseases of current interest. It is for this reason that unaffected family members are of less importance for determining or disproving linkage than affecteds. However, their marker information may be invaluable in defining the segregation within the family (for instance, if one of the parents is not available for typing).

The most informative individuals in a linkage study are those who are affected or unaffected but with a high risk of disease. In the middle of the penetrance range (e.g. a risk of 0.2–0.8), minor changes in risk will have limited impact on the linkage analysis results. If there are groups with very high or very low risks, then liability classes should be maintained for those individuals. In the middle of the range, risks that are discrepant by the order of 0.2 have limited effect on the results and so these groups can be considered as a single liability class.

2.3 Genetic heterogeneity

Heterogeneity is a common concern in genetic analysis. There are several forms of heterogeneity, which cause differing degrees of concern and have different implications for genetic analysis. ‘Phenotypic heterogeneity’ means that disease expression between families is variable so that there may be subtle, or even quite major, differences in the disease expression; ‘linkage heterogeneity’ means that different loci may give rise to the same phenotype, as is the case for a number of syndromes such as retinitis pigmentosa and tuberous sclerosis: in these families there are therefore differing recombination fractions between disease and markers in different families [12]. To a varying degree there are solutions to each of these problems.

2.3.1 Phenotypic heterogeneity

Observed systematic variation in disease expression between families with apparently the same disease may be due to linkage heterogeneity or allelic heterogeneity – that is, where different alleles at the same locus produce differing disease expression. It is important to note at this stage that phenotypic heterogeneity should refer strictly to those instances where disease expression is more consistent within individual families than between families; in fact, an examination of this issue should be the standard

preliminary analysis. If expression is as variable within a family as between families then either the presence of other genes that mediate the expression of the disease gene(s) or interactions with non-genetic factors should be suspected. In this situation, searching for linkage to some 'basic' phenotype is justified followed by examination of other regions as modifiers of the mapped susceptibility. It is worth noting at this time, that incorrectly identifying 'truly unaffected' family members as 'affected' is a considerably more serious error than considering 'affected' individuals as 'unaffected' for the purposes of linkage analysis. In general, therefore, only those with proven disease should be included as affected in these analyses.

There are different ways of handling phenotypic heterogeneity within linkage analysis in different situations; if families are large and can give meaningful lod scores (of 1.0 or more) each, or if there are sufficient families to group them into subsets which have similar phenotypes within each subset and each group can produce a lod score of 3.0 or more, then the solution is straightforward. Simply examine the linkage evidence for each group separately and when one or more of the groups shows significant evidence for linkage then test separately the evidence for linkage of the other groups to that same chromosomal region. Note that the first group to show evidence for linkage must satisfy the usual criterion of a lod score of 3.0 or more, but that subsequent groups do not need to produce such strong support; in these cases, a lod score of the order of 1.00 is sufficient to declare linkage (at an approximate 0.05 significance level), since we are now testing the specific hypothesis that linkage is to that chromosomal region. If one or more of the other groups is, however, unlinked, then a lod score of 3.0 or more will still be required, unless other specific tests can be made (such as testing other candidate regions) on the basis of other information. If families are of limited size and there is variation in phenotype both within and between families, the difficulty of determining linkage will depend entirely on the extent of linkage heterogeneity.

2.3.2 Analysis of linkage heterogeneity

Linkage heterogeneity exists when two or more loci can produce an essentially identical disease outcome. For instance for breast cancer, two high-penetrance genes have been mapped and cloned, *BRCA1* [6] and *BRCA2* [13]. Either of these two genes can produce families which have early onset breast cancer. A typical problem in linkage heterogeneity is

that a single locus has been mapped but it is clear that it does not account for all families with that disease. Furthermore, when only one gene has been mapped, the problem is that the resolution of the precise location of the first gene is difficult in that a family with a potentially informative recombinant can either be classified as a recombinant or as a family which is linked to the other locus. This will not be the case if each family gives overwhelming evidence for or against linkage to the first locus but this is the exceptional situation rather than the rule.

The identification of the location of the first locus and the proportion of families linked to that locus is usually accomplished with 'heterogeneity analysis', in which a statistical model is fitted to the data. The usual way of investigating genetic heterogeneity when a series of families has been collected is to consider a model originally described by Smith but which has been made more popular by Ott (see, for instance, [14]). The model assumes that a proportion of families, α , are due to locus 1 which is at a recombination distance, r , from the marker under consideration. The remainder of families are phenotypically not distinguishable from the linked families but are due to a locus or loci which are unlinked to the first locus. In the analysis, the data presented consist of a lod table for linkage between the marker and the disease for each family taken separately. The result is an estimate of both α and r and the statistical evidence supporting those estimates.

Figure 2.3 shows the result of such an analysis for breast cancer families reported in Easton *et al.* [7] for linkage to *BRCA1*, the gene for familial breast cancer on chromosome 17q. Instead of showing simply the best estimates of α and r , the figure shows the lod scores under the heterogeneity model. The lod scores are plotted for all possible values of α and r ; the magnitude of the lod score shows the relative plausibility of different combinations of α and r with the highest lod scores being the best supported results. The lod score under heterogeneity is plotted on this contour plot so that points on the α and r which have similar lod scores have the same shading. As can be seen from this figure, there are many sets of values of α and r which have similar lod scores and hence are equally persuasive solutions. To fix ideas, consider all those solutions in the lightest shaded part of the surface—that is, those within a lod score of 1.0 of the highest point of the surface. Typically, as in Fig. 2.3, the estimation of α and r is confounded by the lack of knowledge of each; thus, either a smaller proportion of families are linked and the disease locus is closer to the marker

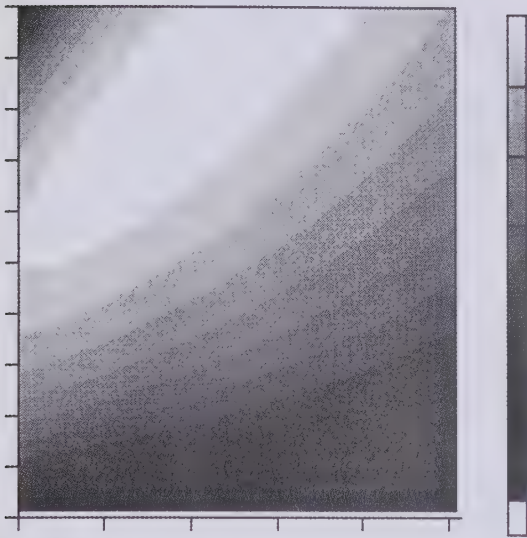


Fig. 2.3 Analysis of linkage heterogeneity for a set of families with breast cancer. These families are slightly modified from those presented in ref. 7. The contour map as a function of the proportion of linked families and the location of the disease gene show the ranges of solutions for these two factors. The data in this case are pairwise lod scores between the disease locus and D17S588 and the maximum lod score is 6.0. All of the unshaded solutions have lod scores of at least 5.5 while those with the lightest shading have lod scores of at least 5.0. If we focus on those with lod scores at most one less than the maximum, then these solutions range for α from 0.37 to 0.73 and for r from 0.0 cM to 25 cM from D17S588.

or a higher proportion are linked but the disease is more distant from the marker.

The analysis of α and r is better achieved by taking multiple markers and basing the analysis on multipoint lod scores. The analysis is most informative when using markers which flank the gene especially when the flanking markers are 10–20 cM apart. Figure 2.4 shows the result of using multipoint analysis as the basis for such a heterogeneity analysis. The confounding of α and r is now no longer so evident and stronger evidence is obtained with better defined estimates; this is particularly true for estimates of α although, as can be seen, there is still some doubt about the location of the disease gene with respect to D17S588. This analysis is typical of the benefit that can be achieved with the use of multipoint analysis for heterogeneity analysis. The resolution of a single locus then helps resolve the location of the second gene by indicating which families are linked to the first locus and hence should not be included in the search for the second. Unfortunately, even with such analyses, if the families are of limited size such resolution will not always be possible and a joint analysis of both loci

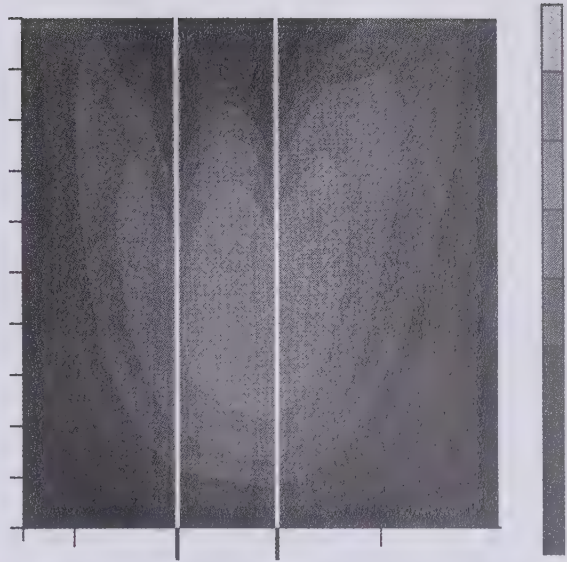


Fig. 2.4 The analysis of heterogeneity based on multipoint analysis involving the breast cancer gene, D17S250 and D17S588. Again the figure shows the relative likelihoods of different solutions for the proportion of linked families and the location of the disease gene. The maximum lod score is 8.1; the lightest shaded region is now smaller than in Fig. 2.3, implying better precision in the estimates.

together with flanking markers (or, putative flanking markers) may be justified.

2.4 Complex traits with no clear mode of inheritance

A number of statistical investigations of the robustness of linkage mapping have come to the same conclusion; that is, for mapping disease genes with major phenotypic effects, misspecification of the mode of inheritance does not invalidate an analysis [15]. This means that misspecifying the model does not lead to an increase in falsely identifying linkage when it is not truly present, while other studies have shown that there is the probability of missing linkage when it is present is also minimally affected [16]. In these analyses, the most deleterious outcome is to assert that the locus has a dominant mode of inheritance when it is in fact recessive, or vice versa. Of course, in general, linkage is most likely to be correctly identified when the correct model is specified [17]. There are therefore important positive features of parametric analysis as shown in Section 2.1 where linkage analysis was shown to take advantage of unaffected individuals; most non-parametric methods consider affected individuals only.

However, many diseases do not show such persuasive evidence for a major Mendelian component and it is therefore less clear that the results concerning power will apply. The approach can still be followed since if a locus is involved in determining susceptibility (although not being the only determinant of susceptibility) then there are reasons to believe that formal linkage analysis could be successful. In this situation, one solution to the concern of incorrectly missing linkage is to examine various modes of inheritance (i.e. a set of assumptions about allele frequencies for the disease susceptibility and penetrance). Because of the concerns of misspecifying dominant as recessive or vice versa, it is natural to try a number of modes of inheritance (i.e. penetrance probabilities) and to look for evidence of linkage with these varied set of penetrance probabilities. The problem is that the usual criteria for identifying linkage (a lod score of 3.0 or more) do not allow for this multiple testing of differing modes of inheritance [18]. Each attempt at a different set of assumptions has a small but definite chance of spuriously identifying linkage. The actual number of different assumptions examined should therefore be limited, and should be specified prior to any analysis being started. A clear conclusion of such studies is that attempting to optimize the lod score leads to a clear accumulation of the type 1 error probability (incorrectly asserting linkage) [17,19].

Non-parametric methods are based on the detection of deviations in the allele-sharing distributions among affected individuals from that expected on the basis of their genetic relationship [20–22]. There are a number of statistical approaches but they are all based on the concept of identity-by-descent—that is, the number of alleles shared by relatives which are direct copies inherited through common ancestors. Sib-pairs may therefore have two alleles identical-by-descent (i.b.d, i.e. they have inherited a copy of exactly the same allele from exactly the same chromosome from each parent as each other), one allele i.b.d. (one allele in common from one parent, the other not), and zero alleles i.b.d. (i.e. for this locus, the chromosomes inherited from each parent were different). In this situation, simple Mendelian genetics shows that the distribution of 2:1:0 alleles i.b.d should be 1:2:1. If in the region of the marker there is a disease gene, then this will be distorted. For instance, if the disease is in fact due to a rare recessive gene, then the allele sharing should all be for two alleles i.b.d. A more complicated situation arises when the i.b.d. sharing cannot be performed exactly because the parents cannot be typed. This situation is termed ‘identity-by-state’

since the fact that two affected siblings share the same allele from a particular parent does not imply that the alleles came from the same parental chromosome [23]. Identity-by-state methods are less powerful than identity-by-descent methods [23].

The appeal of such methods is that affected sib-pairs can be typed for the markers and then a simple statistical test performed which tests for deviations from 1:2:1, the segregation ratio under no linkage. This analysis is not dependent on knowledge of the true mode of inheritance and is therefore more straightforward to apply.

So far in this chapter, we have concentrated on linkage analysis in families as a method of identifying genes involved in disease susceptibility. If, on the other hand, a particular gene is postulated to be involved in susceptibility, then an alternative study design is to compare the allele distribution at this locus in affected and unaffected individuals, and to look for different combinations in the two groups. This approach has been widely used in, for instance, studies of the HLA system, especially in reference to diseases thought to involve an immune response component [24].

A major concern in such studies is that differences between the distribution of alleles in cases and controls are attributable to factors unrelated to the disease process. This would be true if, for instance, the two groups are not matched for geographical location of birth in a situation where the frequencies of the various alleles are not constant over a wide region. So, if cases came predominantly from location A where allele A1 was more common than A2, but controls came predominantly from location B where allele A2 is more common than A1, then there will appear to be a discrepancy in the allele frequencies which is due to geographical variation rather than the disease process itself. Failure to recognize such stratification will produce spurious results. Of course, when the stratification is as simple as that presented above, the problem should be identified, but it may be more subtle, with social structure rather than geographical variation being the cause, and this may not be readily identifiable. To get around these problems, several methods have been developed which rely on sampling more complicated structures than simply affected individuals.

The most straightforward of these methods (but one that may be impracticable with a late-onset disease) is to sample parent–offspring pairs in which the offspring is affected with the disease [25]. Typing both parent and child allows the allele transmitted to the child to be determined, as well as the one not transmitted. More specifically, comparison of the child’s genotype with the combination of alleles not

transmitted to the child permits a comparison in which, by definition, social and geographical stratification is accounted for. In this approach, the genotype produced by combining the non-transmitted alleles forms the ‘control’, although, of course, there is no guarantee that there is a person with such a genotype, or indeed, if one exists, that he or she is unaffected.

2.5 Sampling problems

Some readers may be surprised that the choice of whether to sample extended families or sib-pairs is included as a sampling issue rather than an analysis issue. Unfortunately, the debate is often posed in terms of how the analysis will be performed (i.e. with a non-parametric method or parametric linkage analysis); in fact, once the families have been collected, they can and usually will be analysed in a number of different ways. There is no intrinsic difficulty with this approach (except if carried to an extreme, as many different modes of inheritance are assumed, leading to multiple tests of the same data; see Section 2.3 above).

The most critical issue is to collect family material which is as informative as possible for the linkage study. To be informative in this context, one or more copies of a disease susceptibility allele should be segregating, and as few individuals as possible should be segregating two disease susceptibility alleles. For instance, suppose that for a dominant trait with phenocopies, affected sib-pairs are identified; a proportion of the families will contain two affecteds who are sporadic (i.e. do not carry a copy of the disease allele). For instance, if there is a single susceptibility allele at the disease locus and susceptible individuals are 10 times more likely to

get the disease than non-susceptibles, then Table 2.3 would be true.

Table 2.3 shows the relationship between the frequency of the susceptibility allele (D), the proportion of all cases of that disease in the general population that carry one or more copies of that allele, and the proportion of affected sib-pairs which occur in families segregating D (and hence would be informative in a linkage study). For an efficient analysis, the majority of the sib-pairs should be segregating at least one copy of the disease susceptibility allele (D). Table 2.3 shows that for low allele frequencies of D , sampling affected sib-pairs would not be an efficient way of identifying samples for linkage studies since only a small proportion of all such sib-pairs would be informative (i.e. occur in families in which at least one parent is Dd).

There are few simple solutions to the problem of deciding on the most appropriate sampling unit, and often the major consideration will be availability of families for linkage studies. Unfortunately, the ‘best’ way of collecting families for the study of a specific disease depends upon the true mode of inheritance, which is generally unknown. If extended families are available, then in general these represent the optimal sampling units. The only way in which this becomes a problem is if selecting for such families chooses those that are homozygous for the disease locus, hence reducing the potential informativeness. Unfortunately, it is not possible to exclude this possibility without knowing the true mode of inheritance but it argues for bearing this issue in mind and perhaps choosing samples in different ways to minimize the probability that this problem occurs.

One of the critical issues when planning linkage studies is the number of families that will be

Table 2.3 Relation between the frequency of the susceptibility allele (D) and the proportion of affected sib-pairs that occur in families segregating D .

p	Proportion of cases carrying D	Proportion of sib-pair families segregating D
0.001	0.020	0.108
0.002	0.039	0.196
0.003	0.057	0.268
0.004	0.074	0.328
0.005	0.092	0.379
0.010	0.169	0.550
0.025	0.342	0.743
0.050	0.519	0.820
0.100	0.701	0.830

Assuming that the disease can be due to a dominant gene or to non-genetic factors, then a proportion of all cases will be due to the genetic susceptibility while a proportion of all affected sib-pairs will be segregating for the disease gene, D , and hence are informative in a linkage study. In this table, we suppose that the risk of disease is 10 times increased when carrying at least one copy of D . The population frequency of D is p .

required for an informative study. While this cannot be answered readily in general because it depends upon the true mode of inheritance and the extent of linkage heterogeneity, the minimal sample size which could be informative can be estimated using an approach developed by Risch [26–29]. If a disease is determined by a single gene, then the distortion in the i.b.d.-sharing probabilities is dependent upon the risk of disease in the sibling of a case as compared to the general population. A higher value of this ratio (the ‘relative risk’), the greater the distortion in i.b.d.-sharing induced at that locus, or specifically at a marker locus close to the disease locus. The factors which affect the power of a study are therefore: the relative risk of disease in the siblings, the marker polymorphism and the genetic distance between the disease locus and a marker locus. Thus, for a fully informative marker (i.e. many recognizable alleles) and assuming that the disease and marker are tightly linked and assuming that 60 affected sib-pairs have been collected, the implication is that the power is 80% for a relative risk of 6 or more, 65% for a relative risk of 4.0, 40% for a relative risk of 3.0, 10% for a relative risk of 2.0; this sample size would be acceptable to map a disease gene which was associated with a relative risk of 6.0 or more, but of the order of 250 affected sib-pairs would be required for a relative risk of 2.0. It should be stressed that these are minimal estimates, since they assume that this locus is responsible for all of the relative risk.

Once the families have been collected, typing for a genomic search requires some consideration of the priorities associated with typing specific samples. In general, the most efficient samples are obtained by typing affected individuals in families which give the most convincing evidence of segregating a susceptibility allele. Unaffected relatives should be typed in as far as they give information about the transmission of genes from parent to affected child. Thus, if two parents are available for typing as well as affected and unaffected offspring, in the genomic search it is not worthwhile typing unaffected siblings of the cases. Analysis of the available samples and families can be made through simulation prior to analysis but these methods require knowledge of the mode of inheritance; in terms of ranking the general informativeness of samples, this may be a particularly useful approach.

The initial genomic search is usually conducted with markers that are approximately 20 cM apart; markers closer than this will lead to excessive typing for limited information gain while any more distant will risk missing the disease locus. Of course, such approaches are not foolproof and loci may be

missed. The approach to follow up of any of these linkage results depends then on the number of candidate regions (i.e. the number with lod scores of the order of 3.0) and the number of untyped families (or individuals within families). Typing further families may suffice to include or exclude such loci.

Some further issues related to linkage mapping of complex diseases are discussed in [30].

2.6 Issues of analysis

2.6.1 Looped pedigrees

There are two types of family that lead to complications in performing linkage analysis. These complications are, however, welcome as they imply that particularly informative analyses may well be feasible. A typical problem concerns the investigation of inbred families which contains members with a rare recessive disease; their presence implies that a susceptibility allele is segregating in the family and because the cases are related to each other, the two copies of the disease allele for some or all of the cases may be identical, not only in terms of the precise mutation but also in terms of deriving from the same ancestor. These families are said to contain ‘loops’, requiring modifications to be made to the process by which the likelihood is calculated. The two types are called ‘marriage loops’ (or alternatively ‘exchange’ loops) and inbreeding loops. A typical example of a marriage loop is when two brothers marry two sisters, while an inbreeding loop is created when two related individuals marry—for example, first cousin marriage.

To deal with this issue, LINKAGE requires that we identify a ‘proband’ for each loop. When running MAKEPED, we are prompted with the question ‘DOES YOUR PEDIGREE CONTAIN LOOPS’. At this, answer Y and then insert the number of an individual in that loop. It is appropriate to choose a typed individual where possible, since the analysis will be conducted for each possible genotype of that individual. LINKAGE duplicates the identified individual and associates the ‘original’ and ‘new’ individual with the same marker typing.

For recessive diseases in inbred families, linkage analysis in this setting focuses on identifying regions of homozygosity in a case whose two copies of the mutated allele are likely to be identical and showing that other closely related cases share marker alleles in the same region. In this situation, the region surrounding the important gene is also likely to be homozygous, the regions of heterozygosity surrounding these regions show the boundaries of informative recombination events.

2.6.2 Dealing with large numbers of marker alleles

The large number of alleles at a marker locus has a major impact on the speed of the analysis. The time to compute the lod score for a particular family actually increases with the square of the number of alleles; for instance, an analysis of a system containing 15 alleles will take over twice as long as an analysis involving only 10 alleles. The most noticeable decrease in speed is when there are untyped founder individuals in the family, as is the usual situation for breast cancer families.

There are various ways in which efficiency can be improved.

1 For use in LINKAGE, alleles must be numbered from 1 to *n*, where *n* is the number of different alleles for the marker. For efficiency, therefore, alleles should be numbered consecutively rather than omitting numbers along the way. For instance, for a CA repeat marker with allele sizes from 202 to 220, it might be appealing to number the alleles 2–20 to maintain consistency with the original marker typing, but to improve the speed of analysis recoding the alleles as 1–10 is better.

2 Although there may be a large number of alleles for a particular marker, the number observed for a given pedigree may be much less; see, for example, Fig. 2.5 where only seven alleles are observed. In this situation, improvements can be made by renumbering the observed alleles and considering all ‘unobserved’ alleles as a separate allele. For instance, if the marker in Fig. 2.5 had 11 different alleles, but since only seven alleles are observed in this pedigree, then number the ‘new’ alleles 1–7 and

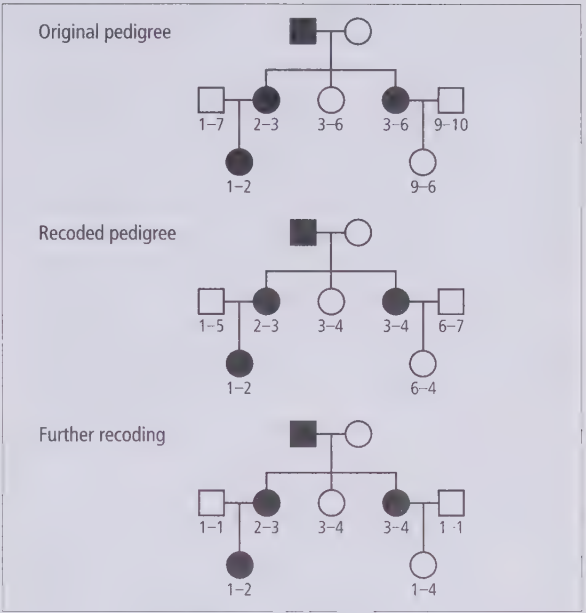


Fig. 2.5 A family for linkage analysis showing the difficulties of using highly polymorphic markers. Untyped individuals add greatly to the time that linkage analysis takes to complete; this is especially true when there are large numbers of alleles at each of the marker loci.

use allele ‘8’ for the remaining unobserved alleles pooled together (see Fig. 2.5 (middle pedigree) and Table 2.4). This requires making a separate DATAFILE (parameter file) for each family or set of families (Table 2.4).

3 Another way of reducing the number of marker alleles required for an analysis is to consider those alleles which only appear a few times in the

Table 2.4 Recoding alleles to improve the efficiency of linkage analysis.

Original allele	Frequency	New allele number	Frequency
<i>Observed alleles</i>			
1	0.1	1	0.1
2	0.2	2	0.2
3	0.05	3	0.05
6	0.05	4	0.05
7	0.1	5	0.1
9	0.01	6	0.01
10	0.15	7	0.15
<i>‘Unobserved’ alleles</i>			
4	0.20	8	0.34
5	0.10		
8	0.02		
11	0.02		

The left columns show the population characteristics of a marker; the right columns the coding for LINKAGE to improve the efficiency of the analysis.

pedigree. Careful examination of their transmission may make it possible to reassign the numbers previously assigned to different allele sizes. One way of doing this is to take advantage of the fact that for dominant diseases (such as breast cancer), the marker alleles transmitted from the non-carrier parent are only relevant in that they define the alleles transmitted by the carrier parent. For instance, in the pedigree in Fig. 2.5, the spouse with the 9–10 marker typing is the only person in the pedigree with either of those alleles. For the linkage analysis, the only information that this person provides is that the 6 allele is transmitted from the mother (the probable carrier); this spouse's genotype could therefore be relabelled in any way that does not blur that information; for example, relabelling that genotype as 1–1 (with the transmitted allele being 1) would retain the linkage information while reducing the number of alleles to be considered in the marker system by 2 (Fig. 2.5 (bottom pedigree)). The basis of this approach is given in ref. 31.

More elaborate recoding is possible, some of which may lose some linkage information. The most general guidance to give is that relabelling should be achieved as much as possible by examining family members who are either not disease mutation carriers or who assist in showing the allele that is transmitted to a disease mutation carrier. One piece of advice is to perform two-point linkage analyses between the disease gene and the marker before and after re-coding (usually two-point analyses can be performed reasonably, the problem is with multi-point analyses) to check that there is no significant loss in linkage information.

2.7 General discussion

This brief introduction to the issues related to complex inheritance can at best serve as a guide to thoughts for such studies. Several basic themes stand out and are appropriate for attempting to map any complex disease: the need for careful evaluation of the number and type of families required, the careful approach to performing the linkage studies and the definition of statistical methodology for the analysis. Other features, however, are more specific to the individual diseases under consideration. Most notably among these issues is that of choice of study design—that is, the choice of materials on which to base predisposition analyses. The choice could be: extended families, sib-pairs or case-control studies looking at associations between candidate genes and disease susceptibility. Several brief comments are in order. If there are extended families available

for study, there are few circumstances in which it is not worthwhile to sample them. The only possible problem with this approach is that it may identify families which do not segregate for the disease gene since the most important family members are homozygous; rarely should this be the case. The next criterion for consideration is the risk of disease in relatives of cases as compared to the general population (the 'relative risk'). If this risk is 3.0 or more and there are numbers of such sib-pairs available, then a sib-pair or similar approach is acceptable. Relative risks of less than 2.0 will require large numbers of sib-pairs (many hundreds under usual conditions) and hence such studies may be prohibitive. The case-control or transmission distortion tests (discussed above) are useful for these more minimal relative risks if candidate genes are postulated. The appropriate choice of study then depends on the parameters of the disease (frequency, relative risk of disease in relatives), the availability of samples (especially sib-pairs) and the knowledge of the disease aetiology which might suggest candidate genes. For this reason, each disease requires careful consideration of its own situation rather than simply applying standard approaches. For instance, various studies [23,27–29] have shown that grandparent-grandchild affected pairs rather than sib-pairs are often the most informative structures for linkage analysis; analysis is then based on whether or not the affected child shares an allele identical-by-descent with the grandparent. Unfortunately, this is usually not a practical design for human studies, especially when dealing with age-dependent diseases, but there may be situations in which it is practical.

References

- 1 Gusella, J.F., Wexler, N.S., Conneally, P.M. *et al.* (1983) A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**, 234–238.
- 2 Bodmer, W.F., Bailey, C.J., Bodmer, J. *et al.* (1987) Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature* **328**, 614–616.
- 3 Leppert, M., Dobbs, M., Scambler, P. *et al.* (1987) The gene for familial polyposis coli maps to the long arm of chromosome 5. *Science* **238**, 1411–1413.
- 4 Beaudet, A., Bowcock, A., Buchwald, M. *et al.* (1986) Linkage of cystic fibrosis to two tightly linked DNA markers: Joint report from a collaborative study. *Am. J. Hum. Genet.* **39**, 681–693.
- 5 Hopper, J.L. & Carlin, J.B. (1992) Familial aggregation of a disease consequent upon correlation between relatives in a risk factor measured on a continuous scale. *Am. J. Epidemiol.* **136** (9), 1138–1147.
- 6 Hall, J.M., Lee, M.K., Newman, B. *et al.* (1990) Linkage of early onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684–1689.

- 7 Easton, D.F., Bishop, D.T., Ford, D., Crockford, G.P. & Breast Cancer Linkage Consortium. (1993) Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. *Am. J. Hum. Genet.* **52** (4), 678–701.
- 8 Claus, E.B., Risch, N. & Thompson, W.D. (1991) Genetic analysis of breast cancer in the cancer and steroid hormone study. *Am. J. Hum. Genet.* **48** (2), 232–242.
- 9 Cox, D.R. & Oakes, D. (1984) *Analysis of Survival Data*. (Monographs on Statistics and Applied Probability 21; Cox, D.R., Hinkley, D.V., Rubin, D. & Silverman, B.W., eds) (Chapman & Hall, Cambridge).
- 10 Mérette, C., King, M.-C. & Ott, J. (1992) Heterogeneity of breast cancer families by using age of onset as a covariate. *Am. J. Hum. Genet.* **50**, 515–519.
- 11 Margaritte, P., Bonaiti-Pellie, C., King, M.-C. & Clerget-Darpoux, F. (1992) Linkage of familial breast cancer to chromosome 17q may not be restricted to early onset disease. *Am. J. Hum. Genet.* **50**, 1231–1234.
- 12 Risch, N. (1989) Linkage detection tests under heterogeneity. *Genet. Epidemiol.* **6**, 473–480.
- 13 Wooster, R., Bignell, G., Lancaster, J. (1995) Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789–792.
- 14 Terwilliger, J.D. & Ott, J. (1994) *Handbook of Human Genetic Linkage* (The Johns Hopkins University Press, Baltimore and London).
- 15 Williamson, J.A. & Amos, C.I. (1990) On the asymptotic behaviour of the estimate of the recombination fraction under the null hypothesis of no linkage when the model is misspecified. *Genet. Epidemiol.* **7**, 309–318.
- 16 Clerget-Darpoux, F., Bonaiti-Pellié, C. & Hochez, J. (1986) Effects of misspecifying genetic parameters in lod score analysis. *Biometrics* **42**, 393–399.
- 17 MacLean, C.J., Bishop, D.T., Sherman, S.L. & Diehl, S.R. (1993) Distribution of lod scores under uncertain mode of inheritance. *Am. J. Hum. Genet.* **52** (2), 354–361.
- 18 Morton, N.E. (1955) Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* **7**, 277–318.
- 19 Weeks, D.E., Lehner, T., Squires-Wheeler, E. *et al.* (1990) Measuring the inflation of the lod score over model parameter values in human linkage analysis. *Genet. Epidemiol.* **7**, 237–243.
- 20 Fisher, R.A. (1935) The detection of linkage with dominant abnormalities. *Annls Eugen.* **6**, 187–201.
- 21 Suarez, B.K., Rice, J. & Reich, T. (1978) The generalized sib pair IBD distribution: its use in the detection of linkage. *Annls Hum. Genet.* **42**, 87–94.
- 22 Holmans, P. (1993) Asymptotic properties of affected-sib-pair linkage analysis. *Am. J. Hum. Genet.* **52**, 362–374.
- 23 Bishop, D.T. & Williamson, J.A. (1990) The power of identity by state methods for linkage analysis. *Am. J. Hum. Genet.* **46**, 254–265.
- 24 Cavalli-Sforza, L.L. & Bodmer, W.F. (1971) *The Genetics of Human Populations* (W.H. Freeman and Co., San Francisco).
- 25 Ewens, W.J. & Spielman, R.S. (1995) The transmission disequilibrium test—history, subdivision and admixture. *Am. J. Hum. Genet.* **57**, 455–464.
- 26 Risch, N. (1990) Linkage strategies for genetically complex traits. I. Multilocus models. *Am. J. Hum. Genet.* **46**, 222–228.
- 27 Risch, N. (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am. J. Hum. Genet.* **46**, 229–241.
- 28 Risch, N. (1990) Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am. J. Hum. Genet.* **46**, 242–253.
- 29 Risch, N. (1992) Corrections to ‘Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs’.
- 30 Risch, N. (1990) Genetic linkage and complex diseases, with special reference to psychiatric disorders. *Genet. Epidemiol.* **7**, 3–16.
- 31 Ott, J. (1978) A simple scheme for the analysis of HLA linkages in pedigrees. *Annls Hum. Genet.* **42**, 255–256.

Chapter 3

Constructing and using genetic maps

Stephen P. Bryant

Gemini Research Ltd, 162 Science Park, Milton Road, Cambridge CB4 4GH, UK

3.1	Introduction, 44	3.3.9	Creating and enhancing reference maps with CRI-MAP, 61
3.2	The principles, 47	3.3.9.1	Constructing data sets for CRI-MAP, 62
3.2.1	Inferring order from recombination information, 47	3.3.9.2	Using CRI-MAP, 63
3.2.2	Likelihood, 48		Protocol 4
3.2.3	Computing probabilities on family data, 48		An illustrative session with CRI-MAP, 64
3.2.4	Interference, 50		
3.3	The protocols, 50	3.3.9.3	Other options not used in Protocol 4, 80
3.3.1	IGD, 52	3.3.9.4	A strategy for using CRI-MAP, 81
3.3.2	X-PED, 53	3.3.9.5	Program notes, 81
3.3.3	CRI-MAP, 53	3.3.10	Genetic maps in disease mapping, 81
3.3.4	Obtaining data, 54	3.3.10.1	Pinpointing a disease gene with LINKMAP, 82
3.3.5	Managing local data, 54		Protocol 5
	Protocol 1		Pinpointing a disease gene with LINKMAP, 83
	Using IGD and X-PED to assist in a linkage analysis, 54		Troubleshooting, 84
3.3.6	Public data within IGD, 56		Typing errors and their detection, 84
3.3.7	Using IGD/X-PED to manage your own data, 57		Problems with the programs, 84
3.3.8	Performing a chromosomal screen with IGD/X-PED and CLODScore, 57	3.4	Integration of physical and genetic maps, 85
	Protocol 2		Acknowledgements, 85
	Creating input files for LINKAGE, 57		References, 86
	Protocol 3		
	Modified procedure for Protocol 2 if a lod score table is required, 60		

3.1 Introduction

Genetic maps show the order and distance between chromosomal landmarks identified by polymorphic markers (Fig. 3.1). They are constructed by following the pattern of marker co-segregation within families, and are an important deliverable of the first phase of the Human Genome Project. The order of loci on a genetic map is statistical, derived using the method of maximum likelihood [1], discussed later in this chapter. The distances between loci on a genetic map are also statistical estimates, and are typically given

in centiMorgans (cM), a unit functionally related to the frequency of recombination by formulae generically referred to as 'mapping functions' [2] (see Chapter 1).

Genetic maps may span anything from a few centiMorgans (corresponding to a few megabases (Mb) of DNA) and contain just three or four markers, to a whole chromosome (200–500cM, or ~200–500Mb) [3] containing several hundred polymorphisms. Sex differences in the rate of recombination mean that the lengths of maps derived from male and female meioses may differ markedly (Fig. 3.1). They form an

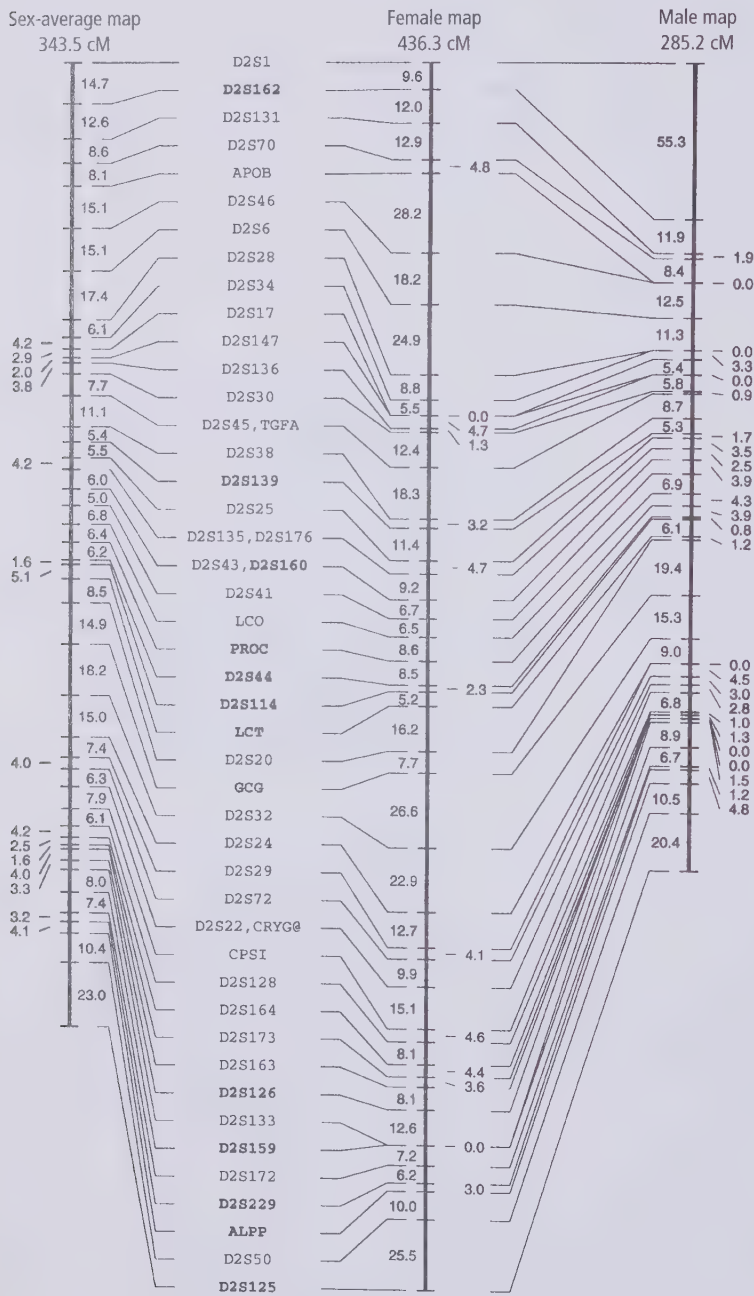


Fig. 3.1 The EUROGEM framework map of human chromosome 2. The order and the recombination fractions between the markers are shown as male, female and sex-averaged. The distances between markers are shown to scale. Loci haplotyped together are shown on the same line. Cytogenetic localization of markers is also shown. This information was obtained from GDB and the localizations were derived using alternative techniques, mainly fluorescent *in situ* hybridization (FISH) or somatic cell hybrid mapping. The markers typed during the EUROGEM project are indicated in bold type. Reproduced with permission of Karger, Basel from [57].

immensely useful resource for the rapid mapping of new DNA markers [4,5] and traits which are influenced by one or several genetic loci spread across the genome [6].

Genetic maps are classified into a number of types, largely by the statistical criteria used to construct them [7]. A framework map is a map where the placement of individual loci have a statistical support of at least 1000:1. This means that the difference in log-likelihoods between the framework map order and any other made by changing the position of any one marker must be at least 3 (since $\log_{10} 1000=3.0$). Alternatively, the support for a map may be tested by permuting the order of markers locally ('flipping') and confirming that the 1000:1 ratio holds for all alternatives. These have also been termed framework maps.

An inclusive map, sometimes referred to as a comprehensive map, is a map where markers are included in their most likely positions irrespective of the statistical support. The utility of such a map is to make statements about the positions of markers which cannot be placed with framework support.

An approximate map is one where the position of markers is shown as the range of intervals which a particular marker could occupy at framework support (Fig. 3.2). These are probably more informative than inclusive maps since the markers do not upset the stability of the framework map.

The discovery of the restriction fragment length polymorphism (RFLP) (see Chapter 5) made it possible to consider constructing maps of the entire human genome using abundant, anonymous DNA markers [8]. At about the same time, it became possible to use personal computers to perform the necessary computations [9,10]. Improvements in algorithms for dealing with large numbers of markers [11] led rapidly to the production of the first high-density map of the human genome [12].

Recent years have seen a rapid increase in the

number of publications of this kind of genetic map. In 1987, the maps published by Donis-Keller and colleagues [12] included 403 markers with an average resolution of 10cM. The maps published in 1994 by the Cooperative Human Linkage Center (CHLC) [13] contained 1123 markers at a resolution of 4.9cM,

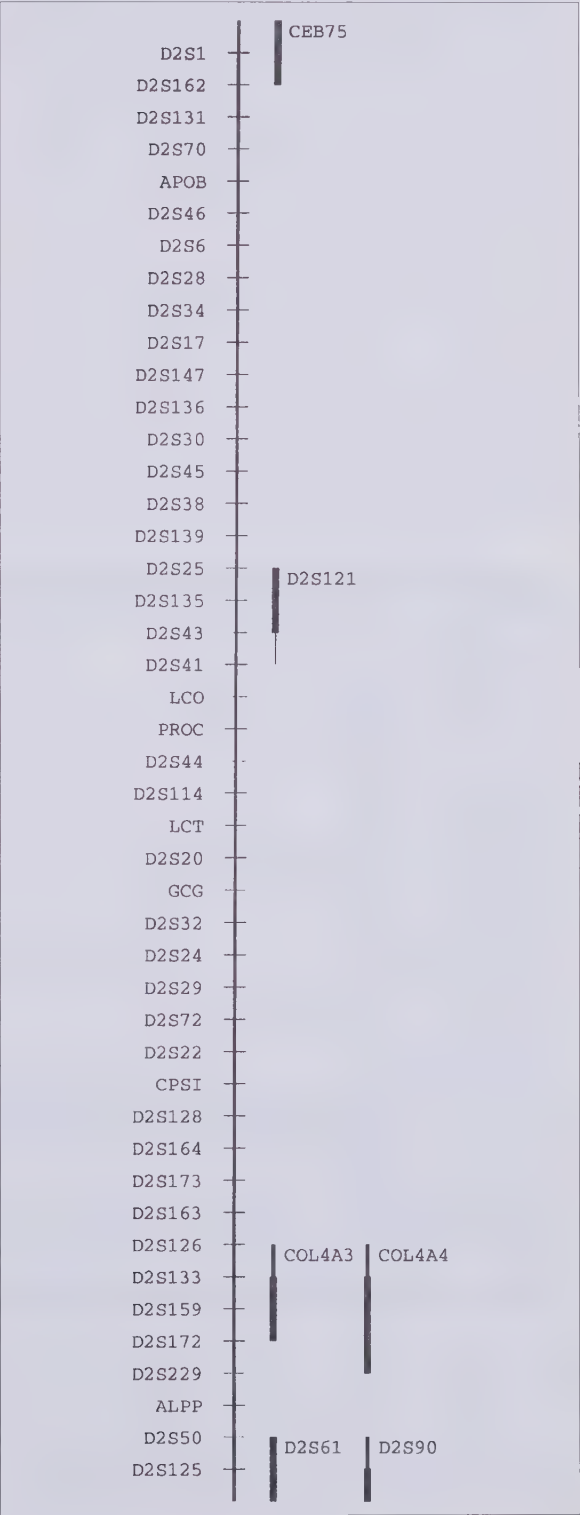


Fig. 3.2 The EUROGEM approximate map of human chromosome 2. This is a simplified representation of the framework map in Fig. 3.1 with markers equally spaced. To the right of the map are indicated the positions of markers which could not be uniquely placed in the framework map. The thickness of the bars indicates the statistical support for each interval. The most likely interval, and others with a log-likelihood difference of less than 1 from it, is shown with a broad line. Intervals with a log-likelihood difference between 1 and 2 compared to the best have a narrower line. Intervals with a log-likelihood difference of between 2 and 3 compared to the best are indicated by a fine line. Reproduced with permission of Karger, Basel from [57].

rapidly followed by a 5840 marker map with a resolution of 0.7 cM [14]. Other notable contributions in this area include the maps from the NIH/CEPH [15] and European Gene Mapping (EUROGEM) [16] consortia. The Centre d'Etude du Polymorphisme Humain (CEPH) consortia have also been a dependable source of maps of specific chromosomes, including 1 [17], 2 [18], 9 [19], 13 [20] and 15q [21]. These maps are usually not now constructed from scratch as has been the case in the past, but instead use existing maps as a starting point for projects aiming to increase their density and overall coverage. There is still plenty of scope for researchers interested in map integration and enhancement (see [22] for an example).

Putting maps of this kind together requires collaboration, both in the sharing of biological resources (markers and DNA) and information (genotypes and maps). One of the best known shared resources of this type is the CEPH [23], based in Paris, under the leadership of Professor Jean Dausset (CEPH, 27 Rue Juliette Dodu, Paris 75010, France; E-mail: cephdbm@ceph.cephb.fr). This resource consists of

DNA, cell lines and genotyping data from 65 two- and three-generation families, containing up to 16 offspring (Fig.3.3). This family structure is very efficient for genetic mapping, in terms of achievable resolution for a certain amount of typing effort. In general, it is not a good idea to use disease families for 'reference' map construction since the amount of information extracted per person will not be optimal and, in addition, it is unusual to have all individuals available for typing, which will significantly complicate the statistical computation and may lead to information being lost. The CEPH works as a network of collaborating laboratories. Each laboratory has a commitment to type new markers across the parents of each family (usually 40, with the other 25 being optional), with a further commitment to type every member of a family where at least one of the parents is heterozygous, and therefore potentially informative for linkage.

When most markers were based on RFLPs [8], this was not a particularly demanding task, since many of the families would be uninformative. But now, microsatellite repeat polymorphisms typically have

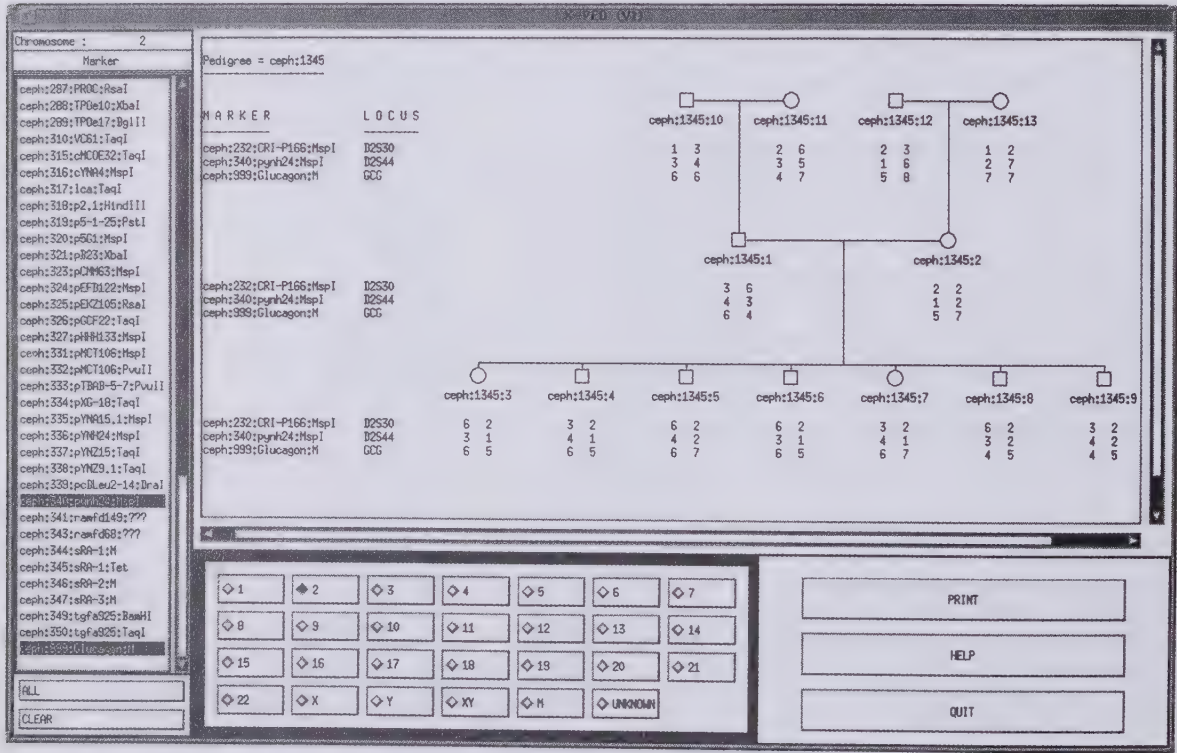


Fig.3.3 CEPH family 1345, displayed using IGD/X-PED. The genotypes for three polymorphic systems on chromosome 2 are shown underneath each person, obtained from the public CEPH database version 7.1. The loci are given in framework order (see Fig.3.1). Paternal alleles are shown on the left of each genotype (for all

individuals with parents in the pedigree) and maternal alleles on the right. The seven cross-overs can be identified by observation. By performing an analysis with CRI-MAP, the relationship between recombinant count and probability can be explored (see Table 3.1).

heterozygosities in the 0.7–0.9 range [24], and the genotyping load is potentially more demanding. What this means, however, is that more information on recombination is being extracted from the CEPH families and that the limit of resolution of maps built with this family panel is being reached. In addition, a great deal of the new data are coming from a small number of highly automated centres such as Généthon [25], which also identify new polymorphisms that can be typed within other initiatives, such as EUROGEM [16].

The CEPH database is publicly available on the World Wide Web (<http://www.cephb.fr>) and is also part of the Integrated Genomic Database (IGD; <http://genome.dkfz-heidelberg.de/igd-docs>). The CEPH families have also formed the basis of the Cooperative Human Linkage Center (CHLC) initiative [26]. This organization has been set up within the United States as a centre of expertise in linkage analysis. The CHLC has developed new markers based on tri-, tetra- and pentanucleotide repeats and has constructed framework maps based on these and other data [13]. The CHLC type their markers across the CEPH families and submit their data to CEPH, making the database an important federation. CEPH does not construct maps by itself, but provides the data for consortia (or individual laboratories) to use. CHLC constructs maps and posts the results on the Internet (<http://ftp.chlc.org>). Other collaborations, such as EUROGEM, do the same (<http://www.icnet.uk>).

3.2 The principles

The most popular kind of polymorphisms in use at the present time are the dinucleotide [24] and trinucleotide [26] repeats (see Chapter 5). These are typically visualized on sequencing gels after PCR, where the genotypes are read directly as the sizes in base pairs (bp) of the alleles. Most loci will be codominant, where each allele is expressed and detectable. Individuals are typed at each locus as a pair of allele identifiers, typically small integers. For two loci, the problem is purely one of estimating the genetic distance (recombination fraction) and significance of the linkage between them, as described in Chapter 1. For three or more loci, there is the additional hypothesis of locus order to consider, as well as distance. Since the number of map orders = $n!/2$, the number of possibilities for a given number of loci rapidly becomes unmanageable.

It is not possible to explore all alternative orders for more than a few loci, nor is it necessary, since the problem can be simplified in several ways. One way is to consider the set of all two-point recombination

fractions and associated lod scores, computed using the program LINKAGE [10]. From these, it is clear which loci are fairly tightly linked (0.05–0.15) with a high confidence (lod score > 3). Using these as a starting point, it is often possible to construct triplets which are ordered with a high degree of support, and simply construct the map as a set of overlapping triplets. This approach has limitations, in that inconsistencies will arise which will need to be dealt with on a statistical basis, and this is where the more sophisticated algorithms become valuable, such as CRI-MAP [27]. This software embodies clever routines which enable a large number of alternative orders to be examined in an heuristic way, avoiding time spent searching poorly supported map orders.

3.2.1 Inferring order from recombination information

Consider the problem of two loci. Here we are concerned solely with the value of a single parameter, the recombination fraction between the two markers, say, S1 and S2. When more markers are added, we can sometimes infer the order from the recombination data. Consider the following example.

Three polymorphic markers from chromosome 2—D2S30, D2S44 and GCG—are typed across a family (Fig. 3.3). The markers are highly heterozygous and all are informative in the family. Recombinants and non-recombinants in the offspring can be scored directly. In this case, the proportion of recombinant chromosomes provides a maximum likelihood estimate of the recombination fraction between any two loci, and hence the genetic distance, between each pair of markers. Changing the hypothesized order of markers will lead to differing numbers of obligatory recombinants (Table 3.1), and may also introduce double recombinants which, because of interference, are very unlikely over short intervals. Now, when this is done, we are interpreting a recombinant as being a single recombinant (and not a triple) and a non-recombinant as being a zero recombinant (and not a double recombinant). If the distance spanned by D2S30, D2S44 and GCG is small, then the

Table 3.1 Recombinant count and log-likelihood affected by marker order, for the family shown in Fig. 3.3.

Order			Recombinants	Log-likelihood
S30	S44	GCG	7	–5.399
GCG	S30	S44	8	–5.461
S30	GCG	S44	10	–6.259

The reader should verify that the two alternative orders introduce double recombinants.

assumption that no double recombinants are possible will hold. In effect, we have assumed complete interference within the region. Then we can exclude those orders which imply double recombinants. Of course, as distances get larger, then double recombinants will appear, and they should be allowed for in the model. If we consider animal models, such as the mouse *Mus musculus*, then it is possible to set up back-crosses (see Chapter 26) so that ordering loci becomes a matter of minimizing the number of recombinants (over short distances). There is a relationship between the minimum recombinant order and the maximum likelihood order (Table 3.1), although, as in much of human genetics, the relationship is not straightforward [28].

Three markers can only be ordered relative to each other if there is at least one recombination between each of them. So, for any data set, there will be a finite number of recombinants to detect. Each recombination event provides information about the placement of markers. Cross-overs bisect the set of markers typed across it. Studies of chiasma counts [29] give the number of observable chiasmata as between two and three per chromosome per meiosis. The only way to increase the resolution of the genetic map, once all cross-overs have been detected, is to increase the number of families typed. The CEPH resource was increased from 40 to 65 families for this reason.

Most of the time, human data are not as ideal as in the above example. One will not know the ‘phase’ of all matings (see Chapter 1), and not all recombinants can be scored explicitly. The process requires the ability to compute multilocus probabilities on pedigrees and make inferences about order and distance between loci. It is here that we need to invoke other kinds of statistical arguments, principally the method of maximum likelihood.

3.2.2 Likelihood

The concept of likelihood can be enshrined in the following statement: ‘The likelihood of a hypothesis, conditional on the observed data, is proportional to the probability of the observed data, conditional on the hypothesis’ [1].

This means that a greater quantitative degree of belief can be put in hypotheses which generate a greater probability for the observed data. Likelihoods tend to be given as exact probabilities, since the constant of proportionality that relates data and hypothesis is usually unknown. Likelihoods have very little meaning by themselves, but when compared (as a likelihood ratio or log-likelihood differ-

ence) they represent our degree of belief in one hypothesis over another. The magnitude of the likelihood ratio (or log-likelihood difference) is used to rank hypotheses and exclude certain of them from further consideration.

In the case of a two-point analysis, the hypothesis of linkage at some value of the recombination fraction is compared against the hypothesis of nonlinkage (free recombination). A likelihood ratio of 1000:1 or a difference in log₁₀ likelihoods of 3 is required to exclude the null hypothesis and demonstrate linkage. The value of the hypothesis (the recombination fraction) that gives the greatest probability to the observations is the Maximum Likelihood Estimate (MLE).

This approach can also be used to compare other kinds of hypotheses, such as the order of loci on a chromosome. Consider the example in Fig. 3.3. If each possible order is taken as an hypothesis, they can be ranked in order of their log-likelihoods (Table 3.1). In this example, the distances between loci are the maximum likelihood estimates. The order which gives the highest probability to the observations is the Maximum Likelihood Order. The alternative orders can be rejected if the differences between their log-likelihoods and that of the best order are large enough. When constructing a framework map, differences of at least 3 are required. In this example, data from the single family are not sufficient to exclude the alternative orders. Further sampling is required to increase the weight of evidence (the support) before a decision can be made.

3.2.3 Computing probabilities on family data

Making maps requires probabilities to be computed on pedigrees, based on some hypotheses about the unknown frequencies of recombination between all the markers.

For a set of phenotypic observations on a pedigree, the exact probability can be expressed as:

$$\sum_{genocoms} \left[\prod_{founders} P(gen) \right] \cdot \left[\prod_{nonfounders} P(gen|pargen) \right] \cdot \left[\prod_{observed} P(phen|gen) \right] \tag{3.1}$$

where:
genocoms = all genotype combinations,
gen = genotype,
phen = phenotype, and
pargen = parental genotype [30].

Equation 3.1 can be applied directly to a single segregating phenotype, but can also be modified to handle more than one locus by incorporating the

recombination fraction(s) as additional parameter(s). With a problem consisting of n loci, the joint estimation of the $n-1$ recombination fractions requires an efficient algorithm which can optimize for many parameters. The Expectation–Maximization (EM) class of algorithms are a very effective way of tackling this problem, and are implemented in the CRI-MAP package [27]. Developments with the algorithm by Lander and Green [11] have meant that the distances between all markers in multilocus maps can be jointly estimated with a small number of iterations (typically less than 10).

The rapid computation of this value for large numbers of loci is essential. Equation 3.1 is rarely implemented directly in the form shown, since the number of permutations rapidly becomes prohibitive. Instead, full likelihood computations are usually based on the algorithm of Elston and Stewart [31] with modifications by Lange and Elston [32] and later by Cannings and others [33] to handle complex pedigrees. The LINKAGE program [10] implements this algorithm and performs a full likelihood computation, whereas the CRI-MAP algorithm computes approximate likelihoods by ignoring data that provide little information and which would be costly to compute, but for the kinds of problems considered here, this information loss is small.

In the model, some parameters may be known and fixed (e.g. allele frequencies), and others may be unknown and allowed to vary within bounds (e.g. recombination fractions). A particular set of hypothetical values will often lead to a different probability conditional on the hypothesized values being true. Whereas absolute probabilities can indicate the MLE of a parameter (or hypothesis), they say nothing about the relative degree of belief which can be associated with that hypothesis compared with any other. A measure of support for a particular hypothesis compared with an alternative can be given as the ratio of two likelihoods, expressed as

$$\frac{L(H_1)}{L(H_2)} \quad (3.2)$$

which provides the odds in favour of H_1 over H_2 . The properties of this ratio are such that the results from different data sets can be multiplicatively combined to provide an overall measure of support. Equation 3.2 is more usually given as the log of the ratio,

$$\log \left(\frac{L(H_1)}{L(H_2)} \right) \quad (3.3)$$

which is equivalent to:

$$\log L(H_1) - \log L(H_2) \quad (3.4)$$

and which can be summed across equivalent data sets. In a linkage analysis, hypotheses which are often compared are those where $H_2: \theta=0.5$ and $H_1: \theta \leq 0.5$ where θ is the recombination fraction between two loci. This measure of support, $Z(\theta)$, where logs are taken to the base 10, is the well-known log-of-odds ratio, or lod score for linkage:

$$Z(\theta) = \log \left(\frac{L(\theta)}{L(0.5)} \right) \quad (3.5)$$

Morton [34] promoted the idea of mapping using lod scores since it offered an elegant solution to the problem of combining data from experiments conducted in different laboratories, even when the primary data were unavailable. Lod scores can be combined from published tables, or from other groups working on different families, until a threshold is reached whereupon linkage is either accepted or rejected. For two autosomal loci, a lod score of 3 is necessary to exclude non-linkage. For two X-linked loci, a lod score of 2 is sufficient. Similarly, a lod score of -2 is sufficient to exclude linkage for a certain distance between two loci. As was hinted previously, several parameters may be estimated simultaneously using this method. These could be a set of recombination fractions in a multilocus map, or the penetrance and allele frequency of a dominant trait. Also, there is no reason why hypotheses should not be discrete, and the method of support is often used to discriminate between different orders of a set of loci.

For example, where there are two alternative orders of a multilocus map, orders which are supported by lod scores of at least 3, against all alternative orders obtained by inverting adjacent loci, are referred to as framework orders, following the definition of framework maps given in Section 3.1.

Another common technique, discussed in detail in Section 3.3.10, is to compute the likelihood at several unknown positions of a marker against a fixed, known map. These plots commonly use \log_e on the ordinate and are referred to as location scores to distinguish them from lod scores. To get the equivalent lod score, divide the location score by 4.6. Logs to the base e have a close relationship to a χ^2 distribution. This technique is an example of the application of reference maps in disease gene mapping.

An alternative to multipoint analysis is the combination of information from multiple two-point analyses. This approach, developed in the MAP package [35] leads to a greatly improved throughput in the map-building process, and is also exploited in the FASTMAP program of Curtis and Gurling [36]. These algorithms are not considered further in this

chapter and the reader is referred to recent reviews on the subject [37–39].

3.2.4 Interference

A complicating factor with human data is that recombination events are not independent and instead interfere with each other [40], so that the relationship between recombination fraction and genetic distance is non-linear. This means that recombination events tend to space out along a chromosome. If interference were zero, then recombination frequencies could be represented by a

Poisson-type distribution. If interference were complete, then chromosomes would have at most one cross-over, and ordering would be a matter of minimizing the number of recombinants in the data set, as discussed earlier.

3.3 The protocols

This section discusses the protocols that are needed to proceed from genotyping data to a well-supported genetic map. It covers the ways of obtaining public data and converting data formats to those used by the analytical programs. The examples, while small,

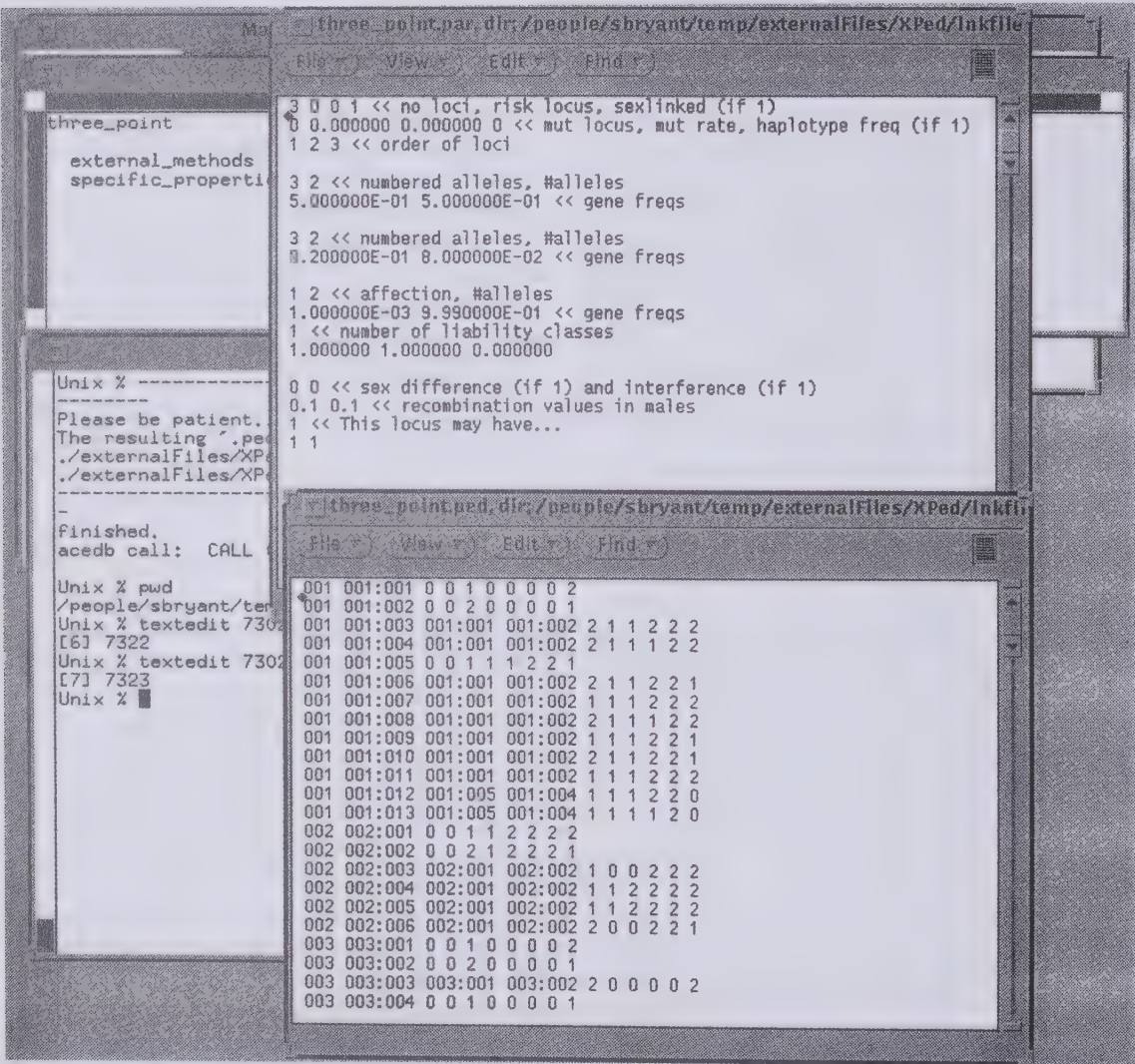


Fig.3.4 A pair of LINKAGE files created using IGD/X-PED, illustrating the LINKAGE data format. The parameter file (top) provides information on each locus; the number of alleles, their frequencies and, for disease loci, the transmission model. Comments in the file appear after the << mark. The pedigree file (bottom) has one line for each individual giving, in order, the identifiers for the

family, the person, the father and mother (0 if founder) and the genotypes (for codominant loci) and phenotypes (for disease loci). Codominant loci are scored as a pair of integers and disease loci are scored as 1 for unaffected and 2 for affected. The locus order across the page in the pedigree file is the same as the order down the page in the parameter file.

are intended to illustrate the principles used in constructing much larger maps. The UK Human Genome Mapping Project Resource Centre (HGMP-RC) and other comparable organizations provide regular courses in this material and readers are strongly advised to attend such a course before embarking on large-scale analyses. In particular, it should be noted that modifications to the protocols will be required, depending on how the software has been installed.

Throughout the chapter, text printed in plain

courier is typed by the system, text in **courier** is typed by the user. Text enclosed in < > should be substituted with an appropriate value (a password, for example). Text enclosed in [] stands for a keyboard function, most often [return]. The discussion is based on use of a powerful UNIX machine, such as those at the HGMP. Interaction with UNIX and associated programs is given in lower case. Please note that file names in UNIX may be in either upper or lower case, with uppercase being distinct from lowercase; for example, TEST1

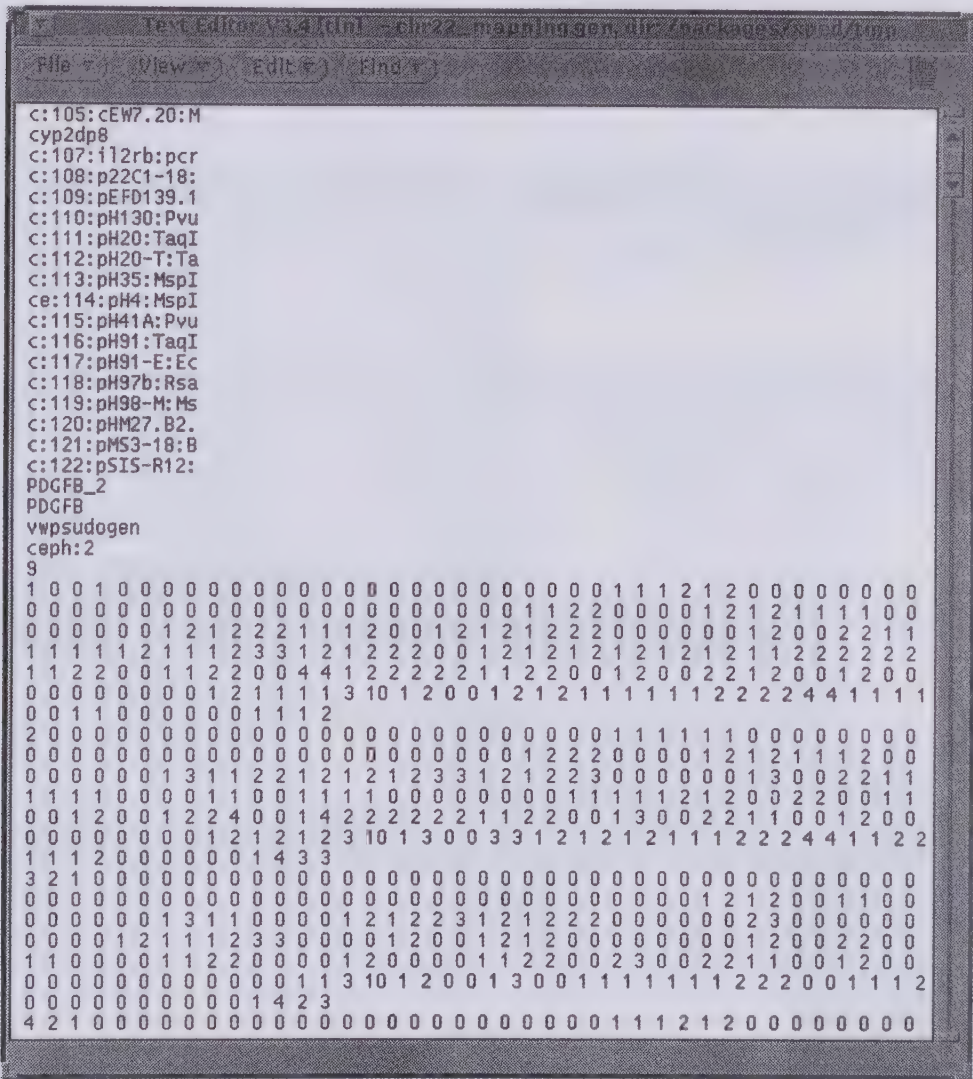


Fig. 3.5 ACRI-MAP genotype (.gen) file created using IGD/X-PED, illustrating the basic format. Data are laid out by family, one person to a line, preceded by a list of the locus names. These names have been derived from the CEPH chromosome 22 database, and have been modified to fit within the 15-character limit demanded by CRI-MAP. Each family has a name (e.g. ceph:2) and an indication of the number of members (e.g. 9). Each person within a family is given a unique id, followed by the ids

of the mother and father (0 if unknown) and a code representing sex (0=female, 1= male). This is followed by pairs of allele numbers for the loci in the order given at the top of the file. The numbers of families and loci are given at the top of the file (off the screen). The file is in 'free' text format, with each item separated from the next by white space (space(s), tab(s) or newline(s)). No comments are allowed anywhere in the file.

and test1 are regarded as *different*. All UNIX files used in this tutorial have lowercase names, which reflects the general preference in UNIX for lower case. For more information on UNIX commands, see Chapter 36.

To sensibly build or improve multipoint maps, one must be able to compute multipoint likelihoods [41]. The most commonly used software for this purpose is CRI-MAP [27]. The most useful function that CRI-MAP performs in map construction is to enable a semi-automatic 'build' process to take place. Managing large data sets for CRI-MAP is facilitated by using database management software specially designed to maintain these kind of data. One good example are the programs produced by the Integrat-

ed Genomic Database (IGD) project [42]. To use publicly available maps in a gene mapping project, the LINKAGE software is used. The rest of this chapter is concerned with the creation, enhancement, integration and use of genetic maps from CEPH reference families using LINKAGE, CRI-MAP and IGD/X-PED.

3.3.1 IGD

The IGD project attempts to integrate data and methods of analysis [42]. This integration is being achieved by writing software that connects other programs together. In current terminology, IGD enables software interoperability. The tool that

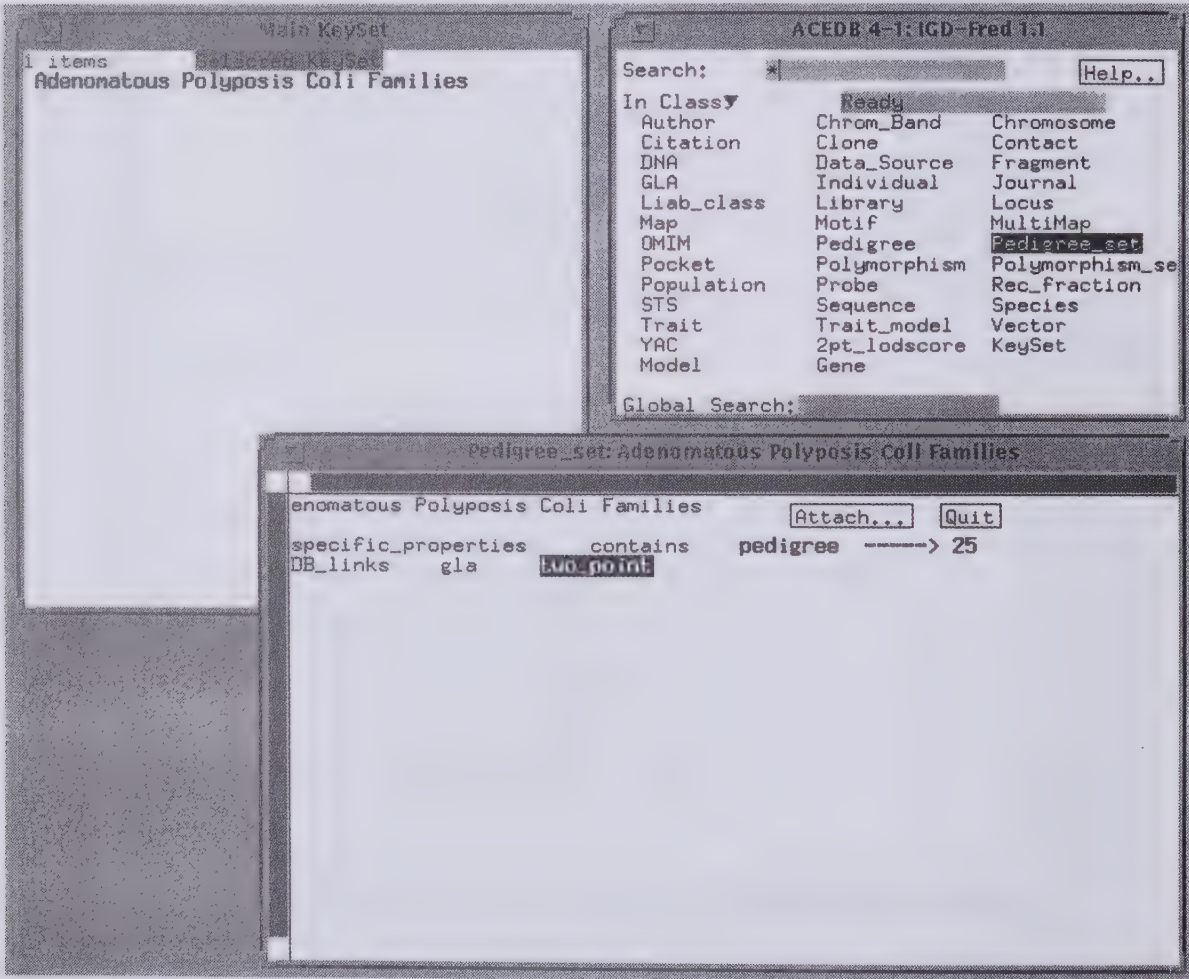


Fig. 3.6 The IGD interface, an implementation of the ACEDB system [43]. Navigation is via hypertext links, similar to those supported by Web browsers such as Netscape. Queries are also supported, as well as the graphical display of complex objects, such as maps and families. In this example, the main window (top right) displays the classes that are available. The user has

selected the Pedigree_set class, which includes only one object (displayed on the left). The object (with name 'Adenomatous Polyposis Coli Families') is shown in the bottom window. The set contains 25 families. Any tag which is highlighted behaves as a 'hypertext' link. IGD works closely with X-PED (Fig. 3.3).

implements much of IGD is the ACEDB system of Richard Durbin and Jean Thierry-Mieg [43]. IGD attempts to construct a federation (Target End Database, or TED) of individual Resource End Databases (REDs), which include the CHLC, CEPH and EUROGEN resources. Data is converted into ACE format, which is presented to the user via the ACEDB software. Extra program modules enable the display and analysis of data not supported directly within ACEDB. IGD programs and data are available on the Web: (<http://genome.dkfz-heidelberg.de/igd-docs>).

3.3.2 X-PED

X-PED (Fig. 3.3 and [44]) works closely in conjunction with ACEDB and the other IGD software components to enable the management and display of pedigree data and the creation of data files for analysis by the LINKAGE and CRI-MAP software packages [45], a function that is examined later. The

implementation includes ‘filter’ programs to convert data into LINKAGE (Fig. 3.4) and CRI-MAP (Fig. 3.5) formats.

3.3.3 CRI-MAP

As indicated earlier, by far the largest problem in generating large maps is in efficiently navigating through the space of all possible orders so as to shorten the search path to a solution that is regarded as optimal. CRI-MAP provides a facility to do this termed ‘build’. ‘build’ uses a set of heuristics in order to decide a sensible way to construct maps.

The current version is 2.4 and is available from Dr Phil Green, Molecular Biotechnology Department, FJ-20, Fluke Hall on Mason Road, University of Washington, Seattle WA 98195, USA (E-mail: phg@u.washington.edu). There is no official ftp server at present. It is distributed as C source code and is particularly easy to port to any C compiler supporting 32 bit (or greater) addressing. It is the

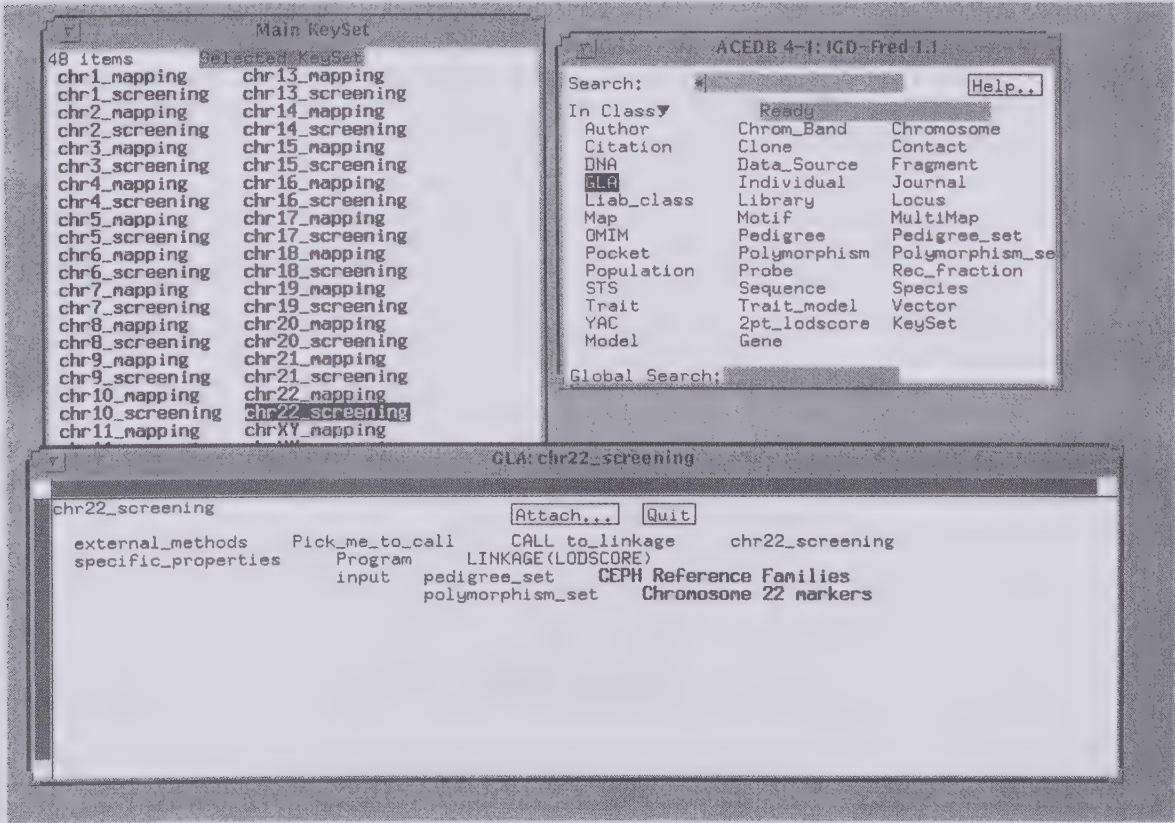


Fig. 3.7 The IGD/X-PED GLA class. The chr22_screening GLA object (which has been set up for screening polymorphisms across the CEPH chromosome 22 database), includes a link to the Pedigree_set ‘CEPH Reference Families’ and the Polymorphism_set ‘

Chromosome 22 markers’. The external method that will be called is to_linkage, which will create LINKAGE files from the data specified in the GLA. As shown, GLA objects have been set up for each chromosome, and can be used as a ‘toolset’ for an analysis of this kind.

software most often used to construct whole chromosome maps [13,16] and is supported at the UK HGMP-RC (<http://www.hgmp.mrc.ac.uk>) and other genome centres. CRI-MAP has been integrated as part of the IGD project with the aim of enhancing the functionality of the program.

Matise and colleagues [46] have developed an expert system (MultiMap) based around CRI-MAP as the engine for computing likelihoods. It builds on the CRI-MAP ability to use heuristics in order to construct a map, and delivers a largely automated system, including error analysis for computing likelihoods. A full description of MultiMap appears in Chapter 4 of this volume.

3.3.4 Obtaining data

Unless you are generating a large amount of data within the laboratory, you will need to obtain some data from a public (or private) repository.

The Cooperative Human Linkage Center (CHLC), which is based in the USA, has put the map and genotype information on the World Wide Web (<http://ftp.chlc.org>). These data are exactly those used to construct the recent genetic maps of the human genome published in *Science* [13]. They are, conveniently, distributed as CRI-MAP format files, but can also be obtained as part of IGD (<http://genome.dkfz-heidelberg.de/igd-docs>).

The CEPH have made their database publicly available on-line (<http://www.cephb.fr>). Data are available as raw genotypes in CEPH ASCII format and may be converted into CRI-MAP format using IGD/X-PED, described later. Version 7.1 has already been converted into ACE format and made available as part of IGD.

3.3.5 Managing local data

Local data can be effectively managed using the IGD/X-PED system (Fig. 3.6). Data are organized into ACEDB classes, which include those for Pedigree, Polymorphism and GLA (which stands for Genetic Linkage Analysis). GLA is a special class which defines the analysis (Fig. 3.7). It creates the association between a set of families (`Pedigree_set`) and a set of polymorphisms (`Polymorphism_set`) along with an indication of the kind of analysis to be conducted (`to_linkage` or `to_crimap` in the `Pick_me_to_call` tag). When mapping disease genes, other classes are also linked to the GLA object, principally `Trait` and `Trait_model`.

Protocol 1 shows how IGD and X-PED can assist in a linkage analysis. It uses the scenario of a session with the UK HGMP-RC (<http://www.hgmp.mrc.ac.uk>) but should be applicable, with slight modifications, to any site implementing the software described.

Protocol 1 Using IGD and X-PED to assist in a linkage analysis

From your workstation, log in to the HGMP-RC. Respond to all the prompts that take you to the main menu. Select the Unix Operating System option and type

```
use xped [return]
```

This command provides access to all the programs we will need in this section (except for LINKAGE, which is shown later). To create an empty database called `Link`, type

```
install_igd_xped Link [return]
```

where `Link` is the name of the directory which will hold the database files.

There are example data sets in the directory

`/packages/xped/example-files` on the HGMP machines. These files should provide all the information you need to set up your own data in IGD/X-PED. The files have the same overall structure (Fig. 3.8). In particular, the structure for the CEPH families is included in `/packages/xped/example-files/ceph_individuals.ace.gz` so that you

will not have to enter this yourself.

Other useful files in this directory are:

`ceph_pedigrees.ace` which sets up the `show_pedigree` tags and assigns the families to a set,

`ceph_contacts.ace` which provides contact information for CEPH collaborators,

`ceph_data_source.ace` which describes the CEPH organization,

`chrn.ace.gz` which contains all the genotypes for chromosome *n* in CEPH version 7.1,

`chrn.set.ace.gz` which assigns the polymorphism for chromosome *n* to a set for ease of use,

`clodscore.gla.ace` which sets up a GLA for each chromosome as a screening method, and

`crimap.gla.ace` which sets up a GLA for each chromosome for map construction.

You should then set ACEDB with

```
setenv ACEDB./Link [return]
```

before calling IGD/X-PED with

```
xace & [return]
```

or else calling IGD/X-PED with

```
xace./Link & [return]
```

The special symbol `&` at the end of the line is the UNIX way of requesting that the command is to be executed in the background, which means that the UNIX prompt will come back after the user hits `[return]`. Within `xace`, selections are almost always made by clicking with the left mouse button. However, with `X`, pull-down menus are accessed by clicking with the right mouse button and holding it down (clicking with the left button would immediately select the first item of the pull-down menu). In some environments you may not need to hold the button down — clicking once on the relevant menu bar item may leave the menu displayed. Items from pull-down menus are chosen by clicking on them. Items from pop-up menus need to be highlighted first and then chosen by clicking on the `[OK]` button. The following exercise illustrates the pertinent features of IGD/X-PED.

When you start IGD/X-PED for the first time as above, you will have no data. `xace` can be used to enter data into the database interactively [43] or data can be imported from files where the data are represented in `ace` format, using the IGD models (see the files in `/packages/xped/example-files` for examples).

You will need either to enter your data interactively using general AceDB editing methods, or create text files in the format of the `chrn.ace.gz` files shown earlier. The public and private data should be combined using the GLA mechanism in IGD. For more information on editing data within ACEDB, see the ACE documentation server (<http://probe.nalusda.gov/acedocs>).

This begs the question: how are the data collected and organized first of all? You would first need to establish conventions for naming polymorphisms and so forth. For the CEPH families, you should use the naming conventions already adopted. You can use text editors (or word

```
//
// Family 5
//

Pedigree : "005"
Pick_me_to_call "CALL show_pedigree" "005"

Pedigree_set : "Adenomatous Polyposis Coli Families"
contains pedigree "005"

Individual : "005:001"
generic_properties local_id "001"
family "005"
sex Female
Phenotype "Adenomatous Polyposis Coli" AFFECTED
genotypes latest "polym2" 1 "polym2:2" "polym2:2"
genotypes latest "polym4" 1 "polym4:1" "polym4:2"
genotypes latest "polym5" 1 "polym5:2" "polym5:2"

Individual : "005:002"
generic_properties local_id "002"
family "005"
sex Male
Phenotype "Adenomatous Polyposis Coli" UNAFFECTED
```

Fig.3.8 An example of the ACE external file format. In the simplest form, ACE data are organized into objects that have a 'class' (e.g. 'Pedigree_set', 'Individual') and a name (e.g. 'Adenomatous Polyposis Coli Families', '005:002'). Each object name can be followed by a series of 'tags' (e.g. 'contains pedigree', 'sex') and associated values (e.g. '005', 'Female').

processors that can save files as plain text) to set up bulk data, or the interactive facilities of IGD/X-PED for small amounts of editing.

To assist, pedigree files in ace format suitable for handling pedigree data have already been set up for use with IGD. In ace format, data from each object is grouped together into a set of lines, one piece of information on each line (Fig. 3.8).

ace files (which conventionally have the suffix .ace) are ordinary text files which can (with care) be modified using an editor such as emacs or xedit. The data files (.wrm) are binary files which live in your ~/Link/database directory and should not be modified directly. The structure of the database and the way in which the interface operates is defined in your ~/Link/wspec directory, and should be modified with great care, if at all.

As you navigate round the database, note how IGD/X-PED lets you move from data item to data item using hypertext-like links that behave in a similar fashion to Netscape and the World Wide Web (Fig. 3.6).

An important part of IGD/X-PED is the ability to display pedigrees and associated genetic data. The way in which this is done is the CALL show_pedigree line in Fig. 3.8. This line allows IGD to call the external X-PED program and pass it the name of the pedigree to be displayed. X-PED extracts the pedigree data from the running database and displays it. You will need to practice this a few times to get used to the X-PED interface (Fig. 3.3).

3.3.6 Public data within IGD

IGD is on the Genome Data option of the HGMP

main menu. To start IGD, select this option. IGD starts automatically and will be most useful in obtaining data that originated in public repositories

such as the CEPH or CHLC. Also included is a large amount of cross-referencing information, from resources such as the Genome Data Base (GDB).

One of the most important issues when dealing with database integration is to make sure that data can be exchanged between the variety of formats needed. We will see how IGD and some other utility programs can help in this process.

3.3.7 Using IGD/X-PED to manage your own data

If you wish to use IGD/X-PED to manage your own data, and analyse these in conjunction with public data, the following will be of interest.

I will assume that you have the set of CEPH pedigrees in a pictorial or other paper-based form, on which have been recorded genotypes (polymorphism information) as a pair of allele codes.

The first thing you will have to do is create an empty database, using the basic IGD structure. From your lab workstation (which must be able to run X windows), log in to the HGMP-RC. Respond to all the prompts that take you to the main menu. Select the Unix Operating System option and type

```
use xped [return]
```

This provides access to all the programs you will need. To create an empty database, type

```
install_igd_xped <dbname> [return]
```

where <dbname> is the name of the directory that will hold the database files. You should then set ACEDB with

```
setenv ACEDB ./.<dbname> [return]
```

before calling IGD/X-PED with

```
xace & [return]
```

or else calling it with

```
xace ./.<dbname> &
```

Recall that the special symbol & at the end of the

line is the Unix way of requesting that the command is to be run in the background. For example, to create an empty database called mydb in your HGMP home directory and to start IGD/X-PED using it, you would use the following sequence of commands.

```
install_igd_xped mydb [return]
setenv ACEDB mydb [return]
xace & [return]
```

The other programs you need, LINKAGE and CRIMAP, are both available from the standard HGMP menus. In the near future, IGD/X-PED will also be available as a menu item, and so you will not need to type **use xped** any more. The IGD/X-PED system will be evolving in the near future, and new versions will be posted when they are available. If you need to do anything special with your data (such as dumping and reloading) to take advantage of a new version, this information will be included in the documentation.

3.3.8 Performing a chromosomal screen with IGD/X-PED and CLODSCORE

If you need to map a new polymorphism to a small region, genotyping across a selection of CEPH families followed by a general genome screen is still a reasonable option. Special versions of the LINKAGE programs exist that are tuned for the CEPH families. The procedure is illustrated in Protocol 2 with reference to the CLODSCORE program, a module of LINKAGE, which performs a function analogous to LODSCORE, but is optimized for nuclear families. Data generation from IGD/X-PED is followed by the analysis and interpretation of CLODSCORE output. If it is necessary to construct a lod score table, the MLINK program can be used, since a CEPH version of this program does not exist.

Protocol 2 Creating input files for LINKAGE

This introduces the LINKAGE package and the way in which it is interfaced to IGD. IGD acts as a shell to the LINKAGE program by creating input files in the appropriate format. The CLODSCORE module of LINKAGE can perform maximum likelihood estimation of the recombination fraction between two polymorphic markers.

The structure of the LINKAGE input files is important, since occasionally these files need to be edited by hand. LINKAGE organizes information into two files, a parameter file and a pedigree file (Fig. 3.4). The seminal

text for practical LINKAGE documentation is the book by Terwilliger and Ott [47].

The rest of this protocol assumes that you have IGD/X-PED running from the previous introductory section, and that you know how to load data from ace files.

IGD/X-PED has been designed specifically to produce data in formats for a variety of different programs (currently LINKAGE and CRI-MAP) which actually carry out the genetic analyses. The way in which sets of pedigrees and polymorphisms are organized to conduct an analysis is the task of the GLA class. To follow this part of the session, you should read `ace file/packages/xped/example_files/ceph/clodscore.gla.ace` in the same way as you read `ceph pedigrees.ace` in the previous exercise. The operation of the GLA class is detailed in the IGD/X-PED documentation and is not repeated here. The problem used as an example here is the mapping of a new polymorphism (D22SXX) thought, as the name suggests, to reside on chromosome 22. The CEPH chromosome 22 data set is used as the resource on which to conduct a screen.

To create the LINKAGE files using the GLA class, select the particular that is required (there should be one called `chr22_mapping`) and double-click on the `Pick_me_to_call` with the `to_linkage` value. This command instructs IGD/X-PED to create the files for LINKAGE. Two files will be created: `nnnn:chr22_mapping.ped` and `nnnn:chr22_mapping.par`, where `nnnn` is a number. The first is the pedigree data file and the second the locus parameter file required by LINKAGE. Both files are initially created in the directory `~/Link/externalFiles/XPed/link-files`, but are also copied to your `~/Link` directory, without the `nnnn` prefix, for convenience. The reasoning behind the use of `nnnn` in the subdirectory is at least partly to avoid overwriting files, and partly to facilitate the parsing of data files back into IGD/X-PED, although this part is not available at present. If you chose a different GLA for the example, all files will include the name of the GLA (here, `chr22_mapping`).

There are two further things that need to be done before the analysis can be performed. The first is to process the `.ped` file into a `.ppd` file. This is achieved using the `makeped` program. `makeped` is a utility for preprocessing the pedigree file to include more information on the pedigree structure for use by the algorithms that perform the likelihood calculations. The second task is to set up a Unix shell script which can call the `clodscore` program and save the results of the analysis.

To perform these tasks, it is first necessary to open another X window that will recognize the names of the LINKAGE programs. To do this, go to your HGMP menu window and navigate to the LINKAGE option on the General Linkage menu which is itself part of the Linkage Analysis menu. Select this and you should see a new X window displaying a `Unix %` prompt. It is within this window that you should run the LINKAGE programs. But first check that you are in the `Link` directory (with all the data files), and if not do a `cd ~/Link`.

To transform your `.ped` file, the command needed (in your new LINKAGE window) is the following:

```
makeped chr22_mapping.ped chr22_mapping.ppd n [return]
```

The first argument (**chr22_mapping.ped**) is the name of the input file, the second (**chr22_mapping.ppd**) the name of the output file, and the third (**n**) is a flag to tell **makeped** that your pedigrees do not contain loops (you will be informed by the program if this is not true).

To construct the Unix shell script, we will use the Linkage Control Program, **lcp**. **lcp** sets up temporary data files for a particular analysis and also creates a batch file for the appropriate operating system. **lcp** has a command structure based on the control **[ctrl]** key. The most important commands are **[ctrl-u]** which clears a field and **[ctrl-n]** which moves on to the next screen. The cursor keys move up and down the fields.

Move back to your LINKAGE window, make sure that you are in your Link directory and type

```
lcp [return]
```

When the Input Files screen appears, move down to the Pedigree file name field and type **[ctrl-u]**. Type the name **chr22_mapping.ppd** here then similarly replace the Parameter file name with **chr22_mapping.par**. Replace the Command file name with **chr22_mapping.sh**, the Log file name with **chr22_mapping.out** and the Stream file name with **chr22_mapping.stm**. **[ctrl-n]** to the Pedigree Options screen. Select the Three Generation Pedigrees option and **[ctrl-n]** to the Three-generation Pedigree Analysis Options screen.

Select **CLODScore** and **[ctrl-n]** to the CLODScore—Sex Difference Options screen. With the cursor on No Sex Difference, **[ctrl-n]** to the CLODScore—Locus Specification screen. This is asking for the markers with which to compute the lod scores. There are two sets. LCP will generate code to test all the markers in the first set against all those in the second. Since we need only to check D22SXX against the markers, we can put **D22SXX** in set 1 and all the others in set 2.

Please note carefully that there is a slight complication in that LCP refers to markers as **p1 ... pn** where the order is as given in the parameter file (**chr22_mapping.par**). If there is some uncertainty about the order, keep a hard copy handy. In this case, **p1** (D22SXX) goes into set 1 and **p2 p3 p4 ... pn** (the rest) into set 2. The **[ctrl-o]** key combination is a quick way of getting all the loci from a file into this field.

The starting value of θ is the point at which the iterative algorithm will start. If there is some prior information on the actual MLE of θ , it may be more efficient to use this value rather than the default (0.1). **[ctrl-n]** to the next screen which wraps back. **[ctrl-z]** closes the output file.

Execute the script file **chr22_mapping.sh** by typing

```
chr22_mapping.sh [return]
```

This will take a few minutes to run. Two output files will be created. **chr22_mapping.out** contains reasonably readable output and can be typed on the screen or printed directly. **chr22_mapping.stm** is a stream file and contains unformatted output. It is used as input to the Linkage Report Program (**lrp**) which interprets the data and prepares a readable summary.

When the batch file has finished executing, run

`lrp [return]`

Change the stream file to `chr22_mapping.stm` and hit `[ctrl-n]` at the Input File and Report Title menu and `[ctrl-n]` to select Three-generation Pedigree Reports from the Report Options menu.

Use the cursor keys to select Two-Point Lodscore Report (CLODSCORE) from the Three-generation Pedigree Report Options menu. `[ctrl-n]` to the Two-Point Lodscore Report (CLODSCORE) Formats menu and select Table Format. `[ctrl-n]` to the Report Output Options screen and select Output Report To A File. Hit `[ctrl-n]`, replace the Report file name with `chr22_mapping.txt` and then hit `[ctrl-n]` again to send the results to this file. `lrp` takes a few seconds to lay out the report. `[ctrl-z]` finishes. Examine your report with

`more chr22_mapping.txt [return]`

or by using a text editor. Do your results show linkage of the unknown polymorphism to chromosome 22 (i.e. lod scores greater than 3)?

Protocol 2 shows how to find the maximum likelihood estimate of the recombination fraction between an unknown polymorphism and a set of markers using IGD/X-PED in conjunction with LINKAGE. Protocol 3 shows the modified procedure necessary if a lod score table is required.

.....

Protocol 3

Modified procedure for Protocol 2
if a lod score table is required

Type

`lcp [return]`

When the Input Files screen appears, move down to the Pedigree File Name field and type `[ctrl-u]`. Type the name `chr22_mapping.ppd` here then similarly replace the Parameter file name with `chr22_mapping.par`. Replace the Command file name with `chr22_mapping_tbl.sh`, the Output file name with `chr22_mapping_tbl.out` and the Stream file name with `chr22_mapping_tbl.stm`. `[ctrl-n]` to the Pedigree Options screen. Select the General Pedigrees option and `[ctrl-n]` to the General Pedigree Analysis Options screen.

Select `MLINK` and `[ctrl-n]` to the Test Options screen. Select Multiple Pairwise Lod Table, then `[ctrl-n]` to the Sex Difference Options screen. Select No Sex Difference, then `[ctrl-n]` to the Multiple Pairwise Lod Table Specification screen.

Again, please note carefully that there is a slight complication in that `lcp` refers to markers as `p1 ... pn` where the order is as given in the locus data file (`chr22_mapping.par`). If there is some uncertainty about the order, keep a hardcopy handy. Put `p1` in the First Locus Set field, `[ctrl-o]` in the Second Locus Set field, `.0` in the Recombination Fractions field, and `.01.05.1.2.3.4` in the Other Recomb. field. Type `[ctrl-n]` to the next screen which wraps back. `[ctrl-z]` closes the output file.

Execute the script file `chr22_mapping_tbl.sh` by typing
`chr22_mapping_tbl.sh [return]`

In this particular example, we are running the `mlink` module of LINKAGE. `mlink` will take a few minutes to run. Two output files will be created. `chr22_mapping_tbl.out` contains reasonably readable output and can be typed on the screen or printed directly. `chr22_mapping_tbl.stm` is a stream file and contains unformatted output. It is used as input to the Linkage Report Program (`lrp`) which interprets the data and prepares a readable summary.

When the batch file has finished executing, run
`lrp [return]`

Change the stream file to `chr22_mapping_tbl.stm` and hit `[ctrl-n]` at the Input File and Report Title menu and `[ctrl-n]` to select General Pedigree Reports from the Report Options menu.

Use the cursor keys to select Lod Table Report (MLINK) from the General Pedigree Report Options menu. `[ctrl-n]` to the Lod Table Report (MLINK) Formats menu and select Table Format. `[ctrl-n]` to the next screen and `[ctrl-n]` to the Report Output Options screen and select Output Report To A File. Hit `[ctrl-n]`, replace the Report file name with `chr22_mapping_tbl.txt` and then hit `[ctrl-n]` again to send the results to this file. `lrp` takes a few seconds to lay out the report. `[ctrl-z]` finishes. Examine your report with

`more chr22_mapping_tbl.txt [return]`

or by using a text editor. Note that the LINKAGE programs will only run inside the LINKAGE shell window, and that IGD/X-PED will only run in the Unix shell window after typing `use xped`. You will need to switch between these windows as required.

The tabular form of the MLINK output provides a rough guide as to the shape of the likelihood curve and hence the MLE of θ . Lod scores are summed over all pedigrees. The analysis has made no attempt to differentiate between male and female recombination fractions and instead has assumed that they are equal. To get a more precise estimate with MLINK, increase the number of θ estimates. Compare the tabular form with the MLE values obtained with CLODScore.

.....

3.3.9 Creating and enhancing reference maps with CRI-MAP

CRI-MAP sacrifices the flexibility and comprehensiveness of the LINKAGE programs for speed and unattended map-building using well-characterized codominant markers. Missing data are largely ignored, and no attempt is made to deal with quantitative data, partial penetrance, or any of the other features of LINKAGE. Instead, the program is optimized for the task of building and improving maps of tens or even hundreds of marker loci, particularly

with data from small, fully typed families, such as the CEPH.

CRI-MAP uses a very efficient algorithm for optimizing the likelihood function but does lose some information from potentially uninformative meioses. Population allele frequencies are not used to determine relative phase probabilities in families with untyped founders. In disease families (where often not all members are typed), this could result in almost all the data being lost and emphasizes the need for fully typed reference families. Therefore, likelihoods, lod scores and measures of support will

all reflect the incomplete nature of the analysis and will probably differ from those computed by LINKAGE. However, maximum likelihood orders and distances will be directly comparable.

CRI-MAP can perform several other useful functions apart from searching map orders. It can estimate the recombination fraction and compute lod scores. It can also show where recombinations occur in families. The example Protocol 4 in Section 3.3.9.2 highlights some of these facilities.

3.3.9.1 Constructing data sets for CRI-MAP

Large input files should always be generated automatically, using a program such as IGD/X-PED, as

the chances of transcription errors increase with the size and complexity of the file. IGD/X-PED can construct files for CRI-MAP in a similar way to that shown for LINKAGE in Section 3.3.8. In the session described here, the CRI-MAP files are already prepared, but most of the time you will need to construct CRI-MAP files from either your own or else from public data. An example of the genotype file format is given in Fig. 3.5.

Note that CRI-MAP systems are numbered from zero and that, when constructing the data file from IGD/X-PED, the gene names and D numbers were used as identifiers (and qualified by the addition of _2, _3, etc.) to make them unique. This makes the

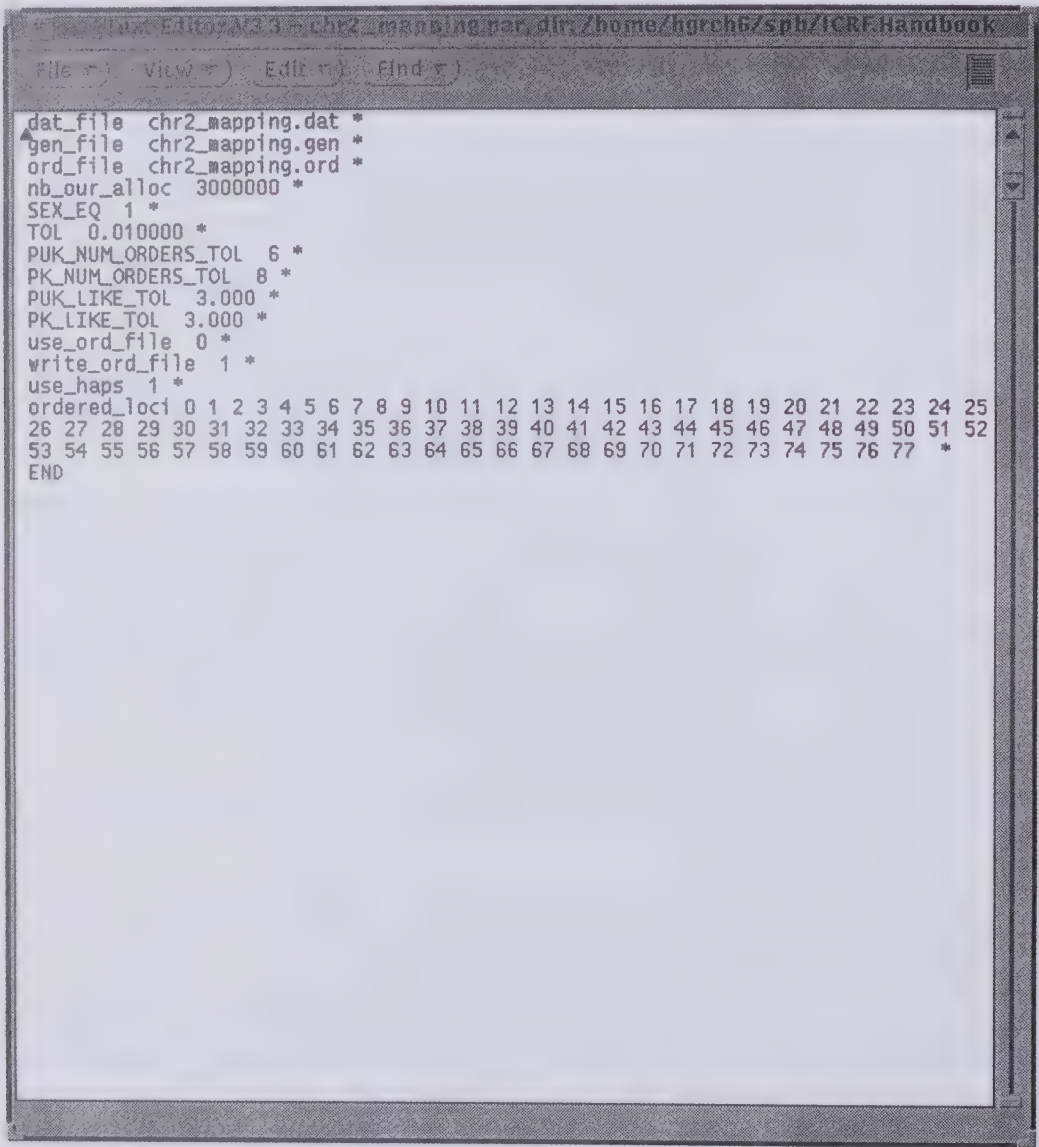


Fig. 3.9 An example of a CRI-MAP parameter file, constructed from the genotype file shown in Fig. 3.5. The name of the genotype file appears on the second line. Underneath are the values for a number of parameters

that control the way in which CRI-MAP behaves, including a line (*ordered_locl*) representing the current best map order (in this case, a default order of all loci).

output from CRI-MAP much easier to read and understand, and it is helpful to know right from the start which systems type the same locus and need to be identified to CRI-MAP as haplotypes.

3.3.9.2 Using CRI-MAP

CRI-MAP is not an interactive program and generates maps in a semiautomatic 'background' mode. The program is always invoked with two command-line parameters: (i) the 'number' of the chromosome to be mapped, and (ii) the program option to be used. Typically, a user types something like

```
crimap 2_mapping build [return]
```

where **crimap** is the name of the executable pro-

gram, **2_mapping** is the 'number' of the chromosome to be mapped and **build** is the name of the option to be run.

In general, CRI-MAP is invoked as

```
crimap <n> <option> [return]
```

where **<n>** refers to a file of the form chr<n>.par and **<option>** can be one of **all**, **build**, **chrompic**, **fixed**, **flips**, **instant**, **quick**, **prepare**, **merge** or **twopoint**. Full descriptions of the options are given in the user guide, supplied with the software. When called in this way (with the exception of the 'merge' and 'prepare' options, described later), the program expects to find a file called chr<n>.par which will contain the names of

Text Editor V3.3 - chr2_mapping.loc, dir /home/hgrchb/sph/ICRF/Handbook

File View Edit Find

▲ Genotype file chr2_mapping.gen

Number of loci: 78

		#inf. mei.	#inf. mei.(phase known)
0	CRYG1-5	343	82
1	D2S54	176	103
2	CRYG1-5_2	352	55
3	D2S54_2	284	62
4	CRYG1-5_3	234	64
5	ACP1	274	130
6	CALMOD	180	55
7	CEB1/HINF	583	439
8	CEB11/HINF	0	0
9	CPSI	251	65
10	D2S25	200	110
11	D2S24	122	48
12	D2S22	160	51
13	D2S36	78	13
14	D2S35	185	73
15	D2S34	158	48
16	D2S21	197	151
17	D2S19	155	39
18	D2S20	139	56
19	D2S42	63	43
20	D2S41	181	50
21	D2S40	120	27
22	D2S39	128	71
23	D2S38	134	62
24	D2S38_2	166	116
25	D2S37	78	53
26	D2S30	217	92
27	D2S32	82	60
28	D2S31	153	59
29	D2S27	139	131
30	D2S29	155	75
31	D2S28	120	90
32	COL5A2	39	26
33	D2S62	210	92
34	D2S16	154	86
35	D2S17	249	58
36	IL1-RN/pcr	43	29
37	IL1A/pcr	41	19
38	IMR-6	119	49
39	D2S5	48	30

Fig. 3.10 An example of a CRI-MAP locus file, constructed from the genotype file shown in Fig. 3.5. It provides the name of the source genotype file and, for each locus, shows how many informative meioses (#inf. mei.) are

present in the data set, along with the number of these that are phase-known. Note that the loci are numbered starting from 0.

the other files in the problem, including the genotype (.gen) file discussed above, which is often given the same prefix (in this example, chr2_mapping.gen). The parameter file is created by CRI-MAP with a special option, 'prepare'. This option must be run prior to any of the others, when only the genotype (.gen) file exists. 'prepare' processes the genotype file into a .dat file and also creates a default.par file, used by the other options.

This is the only interactive part of using the program, and is also an opportunity to set various parameters, such as tolerance and sex-equal analysis, and to group together loci to be treated as haplotypes during future runs. For example, output from the command

```
crimap 2_mapping prepare [return]
```

will be chr2_mapping.dat (the processed data file), chr2_mapping.par (the parameter file), chr2_mapping.loc (the mapping between locus numbers and names, and the numbers of informative meioses for each locus) and, if the 'build' option has been chosen, chr2_mapping.ord (which will become a database of ordering information as the map is built up). The orders database is described more fully in Protocol 4. The processed data file is in an arcane format used directly by the program. However, the parameter file (Fig.3.9) and the locus file (Fig.3.10) are readable and, in particular, the parameter file can be edited with a text editor during

the analysis, which is more productive than interactively altering the file with the 'prepare' option, once created.

Unlike LINKAGE, CRI-MAP has no automatic way of saving its output. It writes everything to the standard output (usually the screen). Therefore, the sensible way to use CRI-MAP is to run it as a background job and redirect its output to a file. A more typical invocation for all but the 'prepare' and 'merge' options would be something like

```
crimap 2_mapping build > mybuild.txt & [return]
```

The 'greater than' sign (>) sends output to the file mybuild.txt instead, and the ampersand (&) causes the job to be run in the background, which gives back the user control of the Unix command line. The file mybuild.txt will grow during the program run and can be examined at will with cat, more, tail or lpr (see Chapter 37). However, most workstations buffer program output in memory, so the output file will remain empty until the buffer fills (say at about 8000 characters) and thereafter grow in chunks of that size. If the program run is aborted prematurely, any output not already saved in the file will be lost. To check the status of the background job, use the Unix jobs or ps commands. When the job disappears from the list displayed, the program has finished.

.....

Protocol 4 An illustrative session with CRI-MAP

Only the first 'prepare' run is shown, for clarity, but a new parameter file, like the one shown in the listing, must be produced before each subsequent CRI-MAP run, either by re-running the 'prepare' option or by editing the previous parameter file with a suitable text editor such as emacs. The example chosen is deliberately small. You will find the data file chr2_mapping.gen in /packages/xped/example-files/chr2 on the HGMP machines. The dialogue can also be applied to your own data. You will need to select the CRI-MAP option from the Linkage menu at the HGMP before you try to run CRI-MAP. In the dialogue, comments are interspersed with the output using this font on a separate line and prefixed with the # character. These will not appear on the screen and are included as a reference. First of all, the 'prepare' option is used, with

```
crimap 2_mapping prepare [return]
chromosome 2_mapping
50400 bytes allocated in orders_morecore
```

```

No.dat file named chr2_mapping.dat
504000 bytes allocated in morecore
Creating.dat file chr2_mapping.dat from.gen file
chr2_mapping.gen
504000 bytes allocated in morecore
504000 bytes allocated in morecore
family id 1326
family id 1327
family id 1328
family id 13281
family id 1329
# Many families have been omitted here for clarity (database errors
# would also be reported here if present).
family id 1477
family id 1582
Writing file chr2_mapping.dat
Finished writing chr2_mapping.dat
Writing locus names to chr2_mapping.loc
Current values for parameters:
par_file = chr2_mapping.par
dat_file = chr2_mapping.dat
gen_file = chr2_mapping.gen
ord_file = chr2_mapping.ord
nb_our_alloc = 3000000
# (Bytes reserved for our_alloc)
SEX_EQ = 1 [0 = sex specific analysis, 1 = sex equal]
TOL = 0.010000
PUK_NUM_ORDERS_TOL = 6
PK_NUM_ORDERS_TOL = 8
PUK_LIKE_TOL = 3.000
PK_LIKE_TOL = 3.000
use_ord_file = 0
write_ord_file = 1
use_haps = 1
Do you wish to change any of these values? (y/n)
y [return]
# Answer yes, since we need to change the tolerance, and switch to
# sex-separate analyses.
To change a value, enter the parameter name, the new value,
and an asterisk, all separated by spaces; for example:
TOL.001 * [return]
Type done when you are finished.
TOL.001 * [return]
SEX_EQ 0 * [return]
done [return]
Current values for parameters:
par_file = chr2_mapping.par
dat_file = chr2_mapping.dat
gen_file = chr2_mapping.gen

```



```

ord_file = chr2_mapping.ord
nb_our_alloc = 3000000
# (Bytes reserved for our_alloc)
SEX_EQ = 0 [0 = sex specific analysis, 1 = sex equal]
TOL = 0.001000
PUK_NUM_ORDERS_TOL = 6
PK_NUM_ORDERS_TOL = 8
PUK_LIKE_TOL = 3.000
PK_LIKE_TOL = 3.000
use_ord_file = 0
write_ord_file = 1
use_haps = 1
The loci and their indices are:
0   CRYG1-5      1 D2S54      2 CRYG1-5_2
3   D2S54_2     4 CRYG1-5_3  5 ACP1
# Many loci have been omitted here for clarity.
72 TGFA      73 TGFA_2    74 D2S90
75 TGFA_3 76 TGFA_4    77 Prot_C/pcr
# We now need to identify the haplotyped systems (systems typing the
# same locus), which we want CRI-MAP to treat as a single unit.
Do you wish to enter any new haplotyped systems? (y/n)
y [return]
For each new haplotyped system which you wish to enter, type either
hap_sys0 (if distances between the loci are to be forced to equal 0)
or
hap_sys (if they aren't),
followed by the indices of the loci to be haplotyped (separated by
spaces), followed by * and a carriage return. Example:
hap_sys 2 0 5 * [return]
When you are done, type
done [return]
To modify or delete a previously entered system, you will need to edit
the .par file later with a text editor.
Ready:
hap_sys0 41 42 * [return] # The two systems typing D2S1.
hap_sys0 39 40 * [return] # The two systems typing D2S5.
hap_sys0 23 24 * [return] # The two systems typing D2S38.
hap_sys0 1 3 * [return] # The two systems typing D2S54.
hap_sys0 53 54 * [return] # The two systems typing APOB.
hap_sys0 0 2 4 * [return] # The three systems typing CRYG1-5.
hap_sys0 72 73 75 76 * [return] # The four systems typing TGFA.
hap_sys0 44 45 * [return] # The two systems typing TPO.
done [return]
Haplotyped system (distances forced to 0.0):
44 TPO      45 TPO_2

```

```

Haplotyped system (distances forced to 0.0):
72 TGFA      73 TGFA_2      75 TGFA_3      76 TGFA_4
Haplotyped system (distances forced to 0.0):
0   CRYG1-5      2 CRYG1-5_2  4 CRYG1-5_3
Haplotyped system (distances forced to 0.0):
53 APOB      54 APOB_2
Haplotyped system (distances forced to 0.0):
1   D2S54  3 D2S54_2
Haplotyped system (distances forced to 0.0):
23 D2S38    24 D2S38_2
Haplotyped system (distances forced to 0.0):
39 D2S5     40 D2S5_2
Haplotyped system (distances forced to 0.0):
41 D2S1     42 D2S1_2
# N.B. Only the first locus in each set is retained in the orders
# objects, but the remaining loci are used in all likelihood
# calculations.
Do you wish to hold any additional recombination frac-
tions fixed? (y/n)
# N.B. These will only be used with the options 'fixed' and 'chrompic',
# and only when the loci in question are adjacent.
n [return]
The crimap options are:
[1] build [2] instant [3] quick [4] fixed
[5] flips [6] all [7] twopoint [8] chrompic
Enter the number of the option you will be running
next: 7 [return]
# Next, we are going to be calculating two-point lod scores with the
# 'two-point' option, so choosing option 7 is correct.
The loci and their indices are:
0   CRYG1-5_1 D2S54      2 CRYG1-5_2
3   D2S54_2      4 CRYG1-5_3  5 ACP1
# Many loci are omitted here for clarity.
72 TGFA      73 TGFA_2      74 D2S90
75 TGFA_3    76 TGFA_4      77 Prot_C/pcr
Do you wish to compute Lod tables for ALL pairs of loci?
(y/n) n [return]
You may specify two separate groups of loci, ordered and inserted. If
both groups are nonempty, 'twopoint' will only compute lod tables for
pairs consisting of one locus from each group. If one group is empty, lod
tables for all pairs from the other group will be computed
Type the indices of the ordered loci (separated by spaces), followed by
a *:
41 50 53 70 56 * [return] # D2S1(41), D2S70(50), APOB(53),
# D2S46(70), D2S48(56).
Ordered loci
41 D2S1      50 D2S70      53 APOB
70 D2S46     56 D2S48

```

Type indices of loci to insert, followed by a *

```
62 * [return]# POMC(62)
```

Inserted loci

```
62 POMC
```

This will cause CRI-MAP to calculate pairwise lod scores between
POMC (locus 62) and D2S1 (41), D2S70 (50), APOB (53), D2S46 (70)
and D2S48. (56). Since we have defined D2S1 as a haplotyped
system, data from systems 41 and 42 will be combined when
calculating lod scores #for this locus, likewise for APOB.

```
OK to set up new parameter file? (y/n)y [return]
```

Always think carefully before answering this, since it overwrites any existing parameter file with the same name.

chr2_mapping.par has been created; use text editor for further modifications, if needed

Take a look at the file with

```
cat chr2_mapping.par [return]
```

```
dat_file chr2_mapping.dat  *# Name of data file to be used.
gen_file chr2_mapping.gen  *# Name of original input file.
ord_file chr2_mapping.ord  *# Name of orders database file.
nb_our_alloc 3000000      *# Allocate memory in chunks of 3 Mb.
SEX_EQ 0 *# Do a sexes-separate analysis.
TOL 0.001000 *# Stop when successive likelihoods differ by <0.001.
PUK_NUM_ORDERS_TOL 6 *# Max phase-unknown orders kept.
PK_NUM_ORDERS_TOL 8 *# Max phase-known orders kept.
PUK_LIKE_TOL 3.000 *# 1000:1 odds for phase-unknown data.
PK_LIKE_TOL 3.000 *# 1000:1 odds for phase-known data.
use_ord_file 0 *# Don't consult orders database ...
write_ord_file 1 *# ... but keep it up-to-date.
use_haps 1 *# Use the haplotypes we defined.
ordered_loci 41 50 53 70 56 *# D2S1, D2S70, APOB, D2S46,
D2S48
inserted_loci 62 *# POMC
hap_sys0 44 45 *# Haplotyped systems as defined.
hap_sys0 72 73 75 76 *
hap_sys0 0 2 4 *
hap_sys0 53 54 *
hap_sys0 1 3 *
hap_sys0 23 24 *
hap_sys0 39 40 *
hap_sys0 41 42 *
END
```

Now we run the analysis with

```
crimap 2_mapping twopoint > pomc.lods &
```

The 'twopoint' option produces a table of lod scores for the chosen combinations of loci. For each pair of markers, CRI-MAP provides the maximum likelihood estimate of θ (optionally separated by sex) and its corresponding lod score, plus a table of lod scores at fixed values of θ (0.001, 0.01, 0.05, and all multiples of 0.05 up to 0.5). When setting up a

'twopoint' run with the 'prepare' option, bear in mind that calculations will be done for each of the 'ordered' loci (`ordered_loci` line) against each of the 'inserted' loci (`inserted_loci` line), but that output will only be produced if the resulting lod score is greater than `PUK_LIKE_TOL`. It is very easy to lose output by leaving `PUK_LIKE_TOL` set too high (or to forget to reset it afterwards and leave too low a threshold during a subsequent build). Note that the tabulated lod scores are not normalized (i.e. not zero at a recombination fraction of 0.5), and need some arithmetic to be performed to make them equivalent to output from, say, MLINK [48]. Now, view the saved results with

```
more pomc.lods [return]
chromosome 2_mapping
50400 bytes allocated in orders_morecore
3024000 bytes allocated in morecore
Option chosen: twopoint
Current values for parameters:
# Here, CRI-MAP repeats the values that we set.
par_file = chr2_mapping.par
dat_file = chr2_mapping.dat
gen_file = chr2_mapping.gen
ord_file = chr2_mapping.ord
nb_our_alloc = 3000000# (Bytes reserved for our_alloc)
SEX_EQ = 0 (0 = sex specific analysis, 1 = sex equal.)
TOL = 0.001000
PUK_NUM_ORDERS_TOL = 6
PK_NUM_ORDERS_TOL = 8
PUK_LIKE_TOL = 3.000
PK_LIKE_TOL = 3.000
use_ord_file = 0
write_ord_file = 0
use_haps =
Haplotyped system (distances forced to 0.0):
41 D2S1      42 D2S1_2
# The remaining haplotyped systems are omitted for clarity.
# N.B. Only the first locus in each set is retained in the orders objects,
# but the remaining loci are used in all likelihood calculations.
D2S1      D2S70      APOB      D2S46      D2S48
AGAINST:
POMC
```

The lod scores follow, showing the best estimates of the recombination fraction in females, then males and the peak lod score in the first line of each entry. The tables of lod scores at recombination fractions 0.001, 0.01, 0.05, 0.10, 0.15 ... 0.5 for females, then males in the following two lines.

```
D2S70 POMC rec. frags.= 0.08 0.00, lods = 3.02
1.51 2.47 2.97 3.01 2.92 2.75 2.54 2.31 2.05 1.79 1.59
1.51
```

```

3.02 3.00 2.91 2.79 2.67 2.54 2.40 2.25 2.09 1.91 1.72
1.52
APOB POMC rec. fracs.= 0.04 0.13, lods = 5.18
3.96 4.87 5.18 4.97 4.60 4.16 3.65 3.10 2.50 1.90 1.35
1.08
0.02 2.95 4.69 5.13 5.17 5.03 4.78 4.44 4.04 3.60 3.18
2.85
D2S46 POMC rec. fracs.= 0.05 0.10, lods =.5.35
4.06 4.98 5.35 5.22 4.94 4.59 4.19 3.76 3.31 2.86 2.44
2.08
2.20 4.13 5.19 5.35 5.24 5.00 4.67 4.29 3.85 3.38 2.91
2.56
D2S48 POMC rec. fracs.= 0.00 0.00, lods = 3.31
3.31 3.30 3.24 3.17 3.10 3.03 2.95 2.88 2.81 2.76 2.72
2.71
3.31 3.27 3.09 2.85 2.61 2.35 2.08 1.80 1.50 1.21 0.90
0.60

```

Note that no result has been given for POMC against D2S1. This is because we left PUK_LIKE_TOL set at 3.0 and only pairs with maximum lod scores greater than or equal to this value are shown in the output. Re-do the parameter file with

```
crimap 2_mapping prepare [return]
```

Continue as previously, but this time set PUK_LIKE_TOL to 0.0. Afterwards, check the file with

```
cat chr2_mapping.par [return]
```

```

dat_file chr2_mapping.dat *
gen_file chr2_mapping.gen *
ord_file chr2_mapping.ord *
nb_our_alloc 3000000 *
SEX_EQ 0 *
TOL 0.001000 *
PUK_NUM_ORDERS_TOL 6 *
PK_NUM_ORDERS_TOL 8 *
PUK_LIKE_TOL 0.000 *# This is what was needed before.
PK_LIKE_TOL 3.000 *
use_ord_file 0 *
write_ord_file 1 *
use_haps 1 *
ordered_loci 41 50 53 70 56 *
inserted_loci 62 *
hap_sys0 41 42 *
hap_sys0 39 40 *
hap_sys0 23 24 *
hap_sys0 1 3 *
hap_sys0 53 54 *
hap_sys0 0 2 4 *
hap_sys0 72 73 75 76 *
hap_sys0 44 45 *
END

```

Try the analysis once again.

crimap 2_mapping twopoint > pomc.lods.again &
and, when the job has finished, view the results.

more pomc.lods.again

The first part has been omitted for clarity.

D2S1 POMC rec. fracs.=. 0.00 0.43, lods = 0.33

All is OK now!

0.33 0.32 0.29 0.25 0.20 0.16 0.13 0.10 0.07 0.05 0.04
0.03

-6.59-3.61-1.58-0.77-0.35-0.08 0.10 0.22 0.29 0.33

0.33 0.30

D2S70 POMC rec. fracs.=.0.08 0.00, lods =.3.02

1.51 2.47 2.97 3.01 2.92 2.75 2.54 2.31 2.05 1.79 1.59

1.51

3.02 3.00 2.91 2.79 2.67 2.54 2.40 2.25 2.09 1.91 1.72

1.52

APOB POMC rec. fracs.= 0.04 0.13, lods =.5.18

3.96 4.87 5.18 4.97 4.60 4.16 3.65 3.10 2.50 1.90 1.35

1.08

0.02 2.95 4.69 5.13 5.17 5.03 4.78 4.44 4.04 3.60 3.18

2.85

D2S46 POMC rec. fracs.= 0.05 0.10, lods = 5.35

4.06 4.98 5.35 5.22 4.94 4.59 4.19 3.76 3.31 2.86 2.44

2.08

2.20 4.13 5.19 5.35 5.24 5.00 4.67 4.29 3.85 3.38 2.91

2.56

D2S48 POMC rec. fracs.= 0.00 0.00, lods = 3.31

3.31 3.30 3.24 3.17 3.10 3.03 2.95 2.88 2.81 2.76 2.72

2.71

3.31 3.27 3.09 2.85 2.61 2.35 2.08 1.80 1.50 1.21 0.90

0.60

Now, try to make a map using 'build'. First, prepare the file.

crimap 2_mapping prepare [return]

Do not use a text editor for this part, since we need CRI-MAP to initialize the order database (.ord file). Start with D2S70 and D2S46 as `ordered_loci` and D2S1, POMC, APOB and D2S48 as `inserted_loci`. Reset `PUK_LIKE_TOL` to 3.0 (since we are building a framework map with 1000:1 odds) and select the 'build' option. Check the parameter file with

cat chr2_mapping.par [return]

`dat_file chr2_mapping.dat *`

`gen_file chr2_mapping.gen *`

`ord_file chr2_mapping.ord *`

`nb_our_alloc 3000000 *`

`SEX_EQ 0 *`

`TOL 0.001000 *`

`PUK_NUM_ORDERS_TOL 6 *`

`PK_NUM_ORDERS_TOL 8 *`

`PUK_LIKE_TOL 3.000 *`


```

PK_LIKE_TOL 3.000 *
use_ord_file 0 *
write_ord_file 1 *
use_haps 1 *
ordered_loci 50 70 *
inserted_loci 41 62 53 56 *
hap_sys0 44 45 *
hap_sys0 72 73 75 76 *
hap_sys0 0 2 4 *
hap_sys0 53 54 *
hap_sys0 1 3 *
hap_sys0 23 24 *
hap_sys0 39 40 *
hap_sys0 41 42 *
END

```

Recall that 'prepare' for the 'build' option creates an orders database (.ord file). Examine this with

```

cat chr2_mapping.ord [return]
1
# The database holds 1 set of orders.
1
# This set has 1 order, of 2 loci.
50 70
# The order is D2S70(50)-D2S46(70).

```

This is the starting point for the 'build' run, one of the most powerful options that CRI-MAP provides. This constructs a map following a set of well-defined heuristics. CRI-MAP can be given a pair of loci, defined as `ordered_loci` in the .par file, and build up as large a map as possible by stepwise addition of the remaining loci (`inserted_loci`). The .ord file is used to keep track of possible 'backtracking' points. For example, if the current map is 1 2 3 and locus 4 cannot be fitted uniquely at the selected level of support but could be either side, say, of locus 2 then the orders database will store two alternative maps, 1 4 2 3 and 1 2 4 3, and the next candidate locus will be tried in both maps. Initially, only phase-known data are used to increase speed, but all the data are used before the locus is finally placed or rejected. The parameter `PK_LIKE_TOL` states, on a \log_{10} scale, by how much the best position must exceed the next best in order for a locus to be provisionally placed. During the subsequent, definitive test, placement is controlled by the value of `PUK_LIKE_TOL`. Normally, both of these would be set to 3 (equivalent to odds of 1000:1).

The .ord file is crucial to CRI-MAP, and it is important to understand how it is used and added to by the programs. Reading and writing the file are controlled by the `use_ord_file` and `write_ord_file` lines in the parameter file. It is made use of only by those options which change orders (e.g. 'build', 'flipsn') and basically remembers where each locus will fit at a given level of support. Thus, if the dataset is too large for a single 'build' run, it is possible to divide it into parts and use the orders database to store the best location for each locus. After this, a single run

with all the loci will very swiftly build a complete map, as the positions of most of the loci are already contained in the database.

However, it is not always desirable to take advantage of the information in the orders database, as it may prevent exploration of other possible orders. For example, running 'flips' on an order generated by 'build' will generate no output if allowed to consult the orders database since the orders database will be consulted before the order of each set of markers is permuted and will almost always rule out the need to try any alternative orders. Care must be taken to use the orders database sensibly, and to keep it in step with the analysis so that it contains correct and up to date information. The orders database will provide enough information to enable most of the analysis to be salvaged in cases where CRIMAP crashes or is terminated prematurely by accidents like power-cuts or accidental reboots. If the file is intact, simply repeating a build run will very quickly recover the best order so far and allow processing to continue from where it was interrupted; otherwise, the quick or instant options can be used to salvage the best map from the information in the database. So, build the map with

```
crimap 2_mapping build > build.no1 & [return]
```

```
[1] 17297
```

Wait a while and examine the results with

```
more build.no1 [return]
```

The preamble has been omitted for clarity

```
41 D2S1
```

```
50 D2S70
```

```
53 APOB
```

```
56 D2S48
```

```
62 POMC
```

```
70 D2S46
```

```
ordered loci:
```

```
50 70
```

```
inserted loci:
```

```
41 62 53 56
```

```
current orders
```

```
50 70
```

This is the starting point for the map (considering phase-known data

only to begin with)

```
current orders
```

```
41 50 70 # Some possible positions for D2S1(41) in the map.
```

```
50 41 70
```

```
50 70 41
```

```
current orders
```

```
50 70
```

```
orders_temp
```

```
41 50 70
```

```
50 41 70
```

```
50 70 41
```

```
orders_temp
```

```
62 50 70
50 62 70
50 70 62
orders_temp
53 50 70
50 53 70
50 70 53
orders_temp
56 50 70
50 56 70
50 70 56
# Nothing definite from phase-known data.
current orders
50 70
# Now start over, using all the data and what we found out so far.
orders_temp
41 50 70
50 41 70
50 70 41
orders_temp
62 50 70
50 62 70
50 70 62
orders_temp
50 53 70
# Got one! APOB(53) fits uniquely now.
current orders
50 53 70
# Record this in the orders database.
orders_temp
41 50 53 70
50 53 70 41
orders_temp
62 50 53 70
50 62 53 70
50 53 62 70
50 53 70 62
orders_temp
56 50 53 70
50 56 53 70
50 53 56 70
50 53 70 56
current orders
41 50 53 70
50 53 70 41
orders_temp
62 50 53 70 41
41 62 50 53 70
50 62 53 70 41
```



```
41 50 62 53 70
50 53 62 70 41
41 50 53 62 70
50 53 70 62 41
41 50 53 70 62
orders_temp
56 50 53 70 41
50 56 53 70 41
41 50 56 53 70
50 53 56 70 41
41 50 53 56 70
50 53 70 56 41
41 50 53 70 56
# Sadly, nothing else goes in at 1000:1 odds.
Sex_averaged map (recomb. frac., Kosambi cM):
50 D2S70 0.0 0.09 9.6
53 APOB 9.6 0.00* 0.0
54 APOB_2 9.6 0.16 17.1
70 D2S46 26.7
* denotes recomb. frac. held fixed in this analysis
# Recombination fractions which are fixed are derived from those
# tems specified as hap_sys0 in the parameter file.
log10_like = -13.98
Sex-specific map (recomb. frac., Kosambi cM - female-
male):
50 D2S70 0.0 0.0 0.06 5.8 0.10 9.9
53 APOB 5.8 9.9 0.00* 0.0 0.00* 0.0
54 APOB_2 5.8 9.9 0.28 32.2 0.00 0.0
70 D2S46 38.0 9.9
* denotes recomb. frac. held fixed in this analysis
log10_like = -10.72
# The best placings of the remaining markers follow at this point.
D2S1
50 53 70
X X
-63.66
# log10(likelihood) for D2S1-D2S70, marginally favoured.
-64.69
# log10(likelihood) for D2S46-D2S1.
POMC
50 53 70
X X X X
-29.20
-28.74
-29.94
-29.14
D2S48
50 53 70
X X X X
```

```

-48.23
-48.41
-47.72
-46.40
# Examine the new orders database with
cat chr2_mapping.ord [return]
2
# The database contains two sets of orders.
7 5
# The first set with 7 maps, each of 5 loci.
56 50 53 70 41
50 56 53 70 41
41 50 56 53 70
50 53 56 70 41
41 50 53 56 70
50 53 70 56 41
41 50 53 70 56
8 5

```

The second set has 8 maps, each of 5 loci.

```

62 50 53 70 41
41 62 50 53 70
50 62 53 70 41
41 50 62 53 70
50 53 62 70 41
41 50 53 62 70
50 53 70 62 41
41 50 53 70 62

```

So, use 'prepare' or a text editor to place our new map (50 53 54 70) as `ordered_loci`, and the others (41 62 56) as `inserted_loci`. We need to test the support for the order by using the 'flips' option.

```
crimap 2_mapping prepare [return]
```

After this is done, check the new parameter file with

```
cat chr2_mapping.par [return]
```

```

dat_file chr2_mapping.dat *
gen_file chr2_mapping.gen *
ord_file chr2_mapping.ord *
nb_our_alloc 3000000 *
SEX_EQ 0 *
TOL 0.001000 *
PUK_NUM_ORDERS_TOL 6 *
PK_NUM_ORDERS_TOL 8 *
PUK_LIKE_TOL 3.000 *
PK_LIKE_TOL 3.000 *
use_ord_file 0 *

```

Must not consult the orders database.

```

write_ord_file 1 *
use_haps 1 *
ordered_loci 50 53 54 70 *
# Our new map.

```

```

inserted_loci 41 62 53 56 *
# 'flips' will ignore these.
hap_sys0 41 42 *
hap_sys0 39 40 *
hap_sys0 23 24 *
hap_sys0 1 3 *
hap_sys0 53 54 *
hap_sys0 0 2 4 *
hap_sys0 72 73 75 76 *
hap_sys0 44 45 *
END

```

Now, run the flip with

```

crimap 2_mapping flips > flips2.no1 & [return]
[1] 17304

```

The 'flipsn' option tests all permutations of n loci down the length of the map (if n is omitted it defaults to 2). It can be used to test the validity of an order generated by other means, flipping 2 or more adjacent loci down the length of the map. For flips ('flips2') all output is given but, for 'flips3' and higher, only those orders with a log likelihood either better than, or within PUK_LIKE_TOL, the starting order are shown in the output. The measure used is $\log_{10}(\text{likelihood ratio})$ and so, because higher likelihoods are closer to 1, better orders show up as negative values. All maps derived from 'build' should be tested with at least 'flips2' (and preferably then with 'flips5') before being regarded as a stable framework. Examine the output with

```

more flips2.no1 [return]
# The preamble has been omitted for clarity.
50 D2S70
# These are the loci in our map.
53 APOB      70 D2S46
number of loci to flip = 2
Original order, & its log10_likelihood, followed by
flipped orders, with their relative log10_likelihoods
(= log10_like[orig] - log10_like[curr])
50 53 70      -10.72
# The starting order.
53 50 -        3.71
# Flip D2S70, APOB —worse by  $\log_{10}$  (5129).
- 70 53        3.49
# Flip APOB, D2S46 —worse by  $\log_{10}$  (3090).

```

So, our new map has local support of greater than 1000:1, and can be termed a framework map. This cycle of 'build' and 'flips' can be iterated as many times as necessary, perhaps starting with a different pair of loci. Of course, there are so many good framework maps that are publicly available for each chromosome, that one would usually be much better off using an existing map as a starting point, in which case the parameter file would be set up with a large `ordered_loci` field, and the whole process becomes one of enhancing the map through inserting and flipping new loci.

During the process of map construction and enhancement, the raw data should be consulted regularly. One of the most useful options, when the map has a reasonably stable order, is 'chrompic'. Using the same parameter file as for the 'flips' run, examine the chromosomes with

```
crimap 2_mapping chrompic > chrompic.no1 & [return]  
[1] 17569
```

The 'chrompic' option is extremely useful to know about as, given an order, it will draw pictures of each child's chromosomes showing the most likely grandparental origin for each locus, and marking the number and position of any cross-overs which have taken place. This can be helpful in spotting errors in the data (e.g. double cross-overs over short genetic distances; too many cross-overs in a single chromosome; most sibs having a cross-over at the same place) and in understanding how the program is interpreting the data. Unfortunately, the cross-overs depend on the order, and vice versa, so this option is of only limited use in the early stages of map development. In interpreting the output it is vital to know that here (and only here) CRI-MAP numbers the loci from their position in the map (starting at 1) rather than their position in the input data file (starting at 0). As elsewhere in CRI-MAP, females precede males, so the top chromosome of the pair is the maternal one, and 0 is used to indicate grandmaternal origin, whilst 1 indicates grandpaternal origin. Digits are used when phase is not in doubt, and letters (i or o) when it cannot be unambiguously deduced. Following the chromosome pictures is an index of individuals with cross-overs between each pair of loci, a list of consecutive markers with no cross-overs between them, and a map of the given order. Examine the output from 'chrompic' with

```
more chrompic.no1 [return]
```

The preamble has been omitted for clarity, and only a few example families shown here.

```
Family 1328 phase likelihood = 0.716, 2d best = 0.284
```

The phase likelihood is a measure of how well it guessed the phase (quite good in this case).

```
3 -ooi      1
```

Cross-over between APOB and D2S46 in maternal chromosome.

```
4 D2S46
```

```
-i-i0
```

No cross-over in paternal chromosome.

```
4 -i-i      0
```

```
-o-o0
```

```
5 -ooo      0
```

```
-o-o0
```

```
6 -ooo      0
```

```
-o-o0
```

```
7 -iio      1
```

```
4 D2S46
```

```
-o-o0
```

```
8 -ii-      0
```

```
-i -        0
```

```

9      -oo-  0
-i -      0
# Two children have been omitted here for clarity.
12 - -      0
- -  0
# All information in this family is phase-unknown (i and o, not 1 and
0), since there are no grandparents in the database.
Family 13291 phase likelihood = 1.000, 2d best = 0.000
# There is no doubt about the phase here, all the data are phase-
known.
1      - -      0
- -  0
2      - -      0
- -  0
3      111-  0
- -  0
4      000-  0
- -  0
5      000-  0
- -  0
6      111-  0
- -  0
7      111-  0
- -  0
8      111-  0
- -  0
9      000-  0
- -  0
14 - -      0
- -  0
Family 13294 phase likelihood = 1.000, 2d best = 0.000
# No real doubt about the phase here—the only cross-over chromo-
some is phase-known.
1 - -      0
- -      0
2      - -      0
- -      0
3      -i1-  0
10 -      1
# Phase-known: D2S70 grandpaternal, APOB grand-maternal.
1 D2S70 2 APOB
4      -o0-  0
0o -      0
5      -o0-  0
11 -      0
6      -i1-  0
1i -      0
7      - -      0
- -      0

```

```
8      -i1-  0
1i  -      0
13  - -    0
-   -      0
14  - -    0
-   -      0

CROSSOVER CHROMOSOMES FOR EACH INFORMATIVE INTERVAL
1      2
# D2S70 [x] APOB
13294-3-P
# Person 3 in family 13294, paternally derived chromosome.
1      3
# D2S70 [x] APOB_2
2-3-M 1354-13-P 1354-12-P
1      4
# D2S70 [x] D2S46
1458-3-M 1454-8-M 1454-7-M 1454-4-M 1454-3-M 1447-3-M
1444-9-M 1444-8-M 1444-4-M 37-8-P
2-5-P 2-4-P 66-7-M 66-4-M 1377-8-M 1377-3-M 1355-11-M
1355-7-M 1345-8-M 1345-5-M
1345-4-M 1340-13-M 1340-5-M 1340-4-M
3      4
# APOB_2 [x] D2S46
1416-8-M 1416-6-M 1344-10-M 1344-9-M 1344-8-M 1328-7-M
1328-3-M
CONSECUTIVE LOCI UNSEPARATED BY CROSSES:
2      3
# The two halves of the APOB haplotyped system.
Sex-specific map (recomb. frac., Kosambi cM - female,
male):
1      D2S70      0.0      0.0  0.06  5.8  0.10 9.9
2      APOB      5.8      9.9  0.00* 0.0      0.00*0.0
3      APOB_2      5.8      9.9  0.28 32.2  0.000 0.0
4      D2S46      38.0      9.9 * denotes recomb.
frac. held fixed in this analysis
log10_like = -10.718
# Overall log10 (likelihood) of the map.

.....
```

3.3.9.3 Other options not used in Protocol 4

‘fixed’ simply calculates the map distances for a given order. ‘quick’ and ‘instant’ are used to construct maps solely from the information already built up inside the the orders database. They do no likelihood calculations whatsoever in constructing the map, but quickly deduce the best map which can be made from the orders in the database. Likelihood

analysis is then used to calculate the intermarker distances in this map and, with the instant option only, the likelihoods associated with the alternative positions for those markers not fitted uniquely into the map. ‘merge’ is called in the usual way (e.g. `crimap 2 merge`) but completely ignores whatever chromosome ‘number’ it is given and asks for the names of

two genotype (.gen) files, which it merges and writes out to a third file. The two files may, if necessary, contain overlapping sets of families and/or loci, and the program will merge the data together, reporting problems along the way. If both files contain data for the same locus in the same individual, then that in the second file will take precedence and overwrite the data in the first. This option is very useful for safely combining data from different sources, something that is difficult (and error-prone) with a text editor. Beware when combining CEPH data from public repositories, that the identifiers for families and individuals are *exactly* the same in the two input files, otherwise CRI-MAP will become terribly confused.

3.3.9.4 A strategy for using CRI-MAP

Best use of CRI-MAP is made by taking advantage of its automatic map-building ability ('build'), but care must be taken to see that the end result is the best map, rather than merely one of a number of possible maps. This means checking the built maps by flipping adjacent loci ('flips'), preferably at several stages during the build process, and running 'build' several times, starting with different pairs of loci (or, later on, with a skeleton of well-ordered loci spanning the area of interest). The two most informative loci (CRI-MAP's default starting point) are not always the best, and alternative pairs may be found by inspecting the table of pairwise lod scores and the list of loci (perhaps sorted by informativeness), or by using other programs with the same data to suggest likely candidates (e.g. CLODSCORE, as described in the previous section). It is not always a good idea to use only one orders database throughout the whole process, as placings in early maps may not turn out to be the best when a more reliable starting point has been found. However, it is always worth keeping orders databases built up during the map-building process, as they may contain information which can be exploited in later runs. Orders files can be kept by copying them with the Unix cp command:

```
cp chr2_mapping.or dchr2_mapping.ord
.safe [return]
```

To stop CRI-MAP from using the orders file, or to make it use one of the saved ones, edit the .par file with a text editor, such as xedit. Follow the 'build' and 'flips' runs with a series of 'all' runs, which will place loci in their approximate positions, bounded by a support interval defined by PUK_LIKE_TOL.

3.3.9.5 Program notes

All likelihood calculations in CRI-MAP are performed iteratively, starting from an initial first guess, which is improved by an equation (the layered EM

algorithm) and then fed back through the equation until two successive answers differ by less than a specified amount (called the tolerance). At this point, the current answer is considered to be sufficiently close to the correct answer, and the calculation is terminated. Accuracy can therefore be increased (at the expense of speed) by decreasing the tolerance. The tolerance parameter in the parameter file, which defaults to 0.01, is better when set to 0.001. This can be set once at the beginning of the analysis and will be propagated, via the parameter file, throughout subsequent runs.

If care is exercised, parameter files can be changed with a text editor once the original run has been set up with the 'prepare' option. However, if a completely new order is to be used, the parameter file should be recreated in order to re-initialize the orders database too. After a successful build, the new order should if possible be cut and pasted from the output file into the parameter file to avoid transcription errors.

All output from CRI-MAP is given as \log_{10} (likelihoods). Likelihoods themselves are constrained within the range of 0.0 (impossible) to 1.0 (certain). Since $\log_{10}(0.0)$ is minus infinity, and $\log_{10}(1.0)$ is 0.0, this means that a \log_{10} (likelihood) of -256 is more likely than a \log_{10} (likelihood) of -260. As described in Section 3.2.3, the absolute value is meaningless, but the difference between the two (4) is a measure of the odds in favour of one over the other (10000:1 in this case). \log_{10} (likelihoods) may only be compared when they are derived from the same data, under different hypotheses, such as alternative orders of the same set of markers, or sex differences in recombination rates for a given map. Thus, if the \log_{10} (likelihood) of the order 1-2-3-4 is -100, and the \log_{10} (likelihood) of an alternative order 1-3-2-4 is -103, then we can say that the first order is more likely to be correct (it has the higher likelihood) and that the odds favour this order over the alternative by 1000:1 ($\log_{10}(103-100)$). However, if the \log_{10} (likelihood) of the order 1-3-2 is -92, we cannot say that it is more likely than the order 1-2-3-4, because they are not comparable hypotheses. In mapping, a \log_{10} (likelihood) difference of 3 (or odds of 1000:1) is generally taken as the minimum measure of significance.

3.3.10 Genetic maps in disease mapping

Genetic maps such as those discussed in this chapter find their greatest utility as tools for mapping disease genes. Recently, Davies and colleagues [6] used genetic maps to conduct a genome-wide scan for susceptibility loci for insulin-dependent diabetes

mellitus (IDDM). They used conventional multi-point methods to construct the map, and an affected-sib method to look for significant allele sharing in their disease families. This method could be applied generally to screen for genetic effects, and maps with a uniform high density of highly polymorphic markers then become essential.

Reference maps also provide input into a LINKMAP analysis, where the position of an unknown trait locus is tested against a fixed reference map. The distances between the markers on the map are computed from reference families such as the CEPH and not in the disease families themselves. Protocol 5 shows how to use a reference map in a LINKMAP analysis.

3.3.10.1 Pinpointing a disease gene with LINKMAP
A multipoint marker map, obtained using CRI-MAP or taken from a public repository such as IGD, could effectively be used as a template on which to place a disease locus. The statistical ideas are very similar to estimating the recombination fraction between two loci. The competing hypotheses are now the position of the unknown disease susceptibility locus relative to a fixed map of marker loci. The most likely position of the unknown locus is that which maximizes the likelihood of the resulting map (and the probability of the observed data). In this exercise, the

LINKMAP component of LINKAGE is used to place a locus predisposing to adenomatous polyposis coli (APC). APC has been previously mapped to chromosome 5 by linkage [49], and the example data set is from that paper. The data for the exercise are available on the Web, but the protocol is generally applicable to any comparable data set.

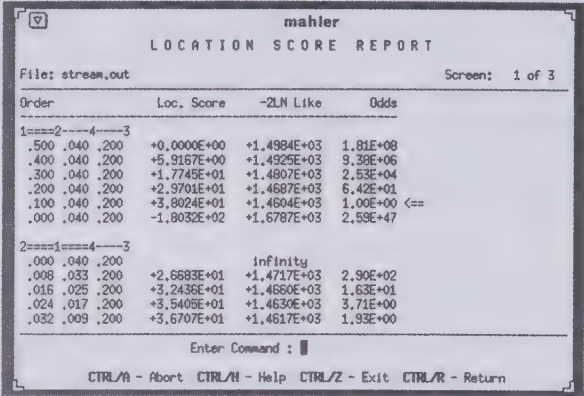


Fig. 3.11 A screen from LRP showing part of the results of a LINKMAP analysis. LRP will highlight the interval within which the maximum likelihood location lies. Subsequent screens show alternative intervals with the odds ratio given against the most likely interval. A ratio of greater than 1000:1 is deemed to be significant. Reproduced with permission from [38].

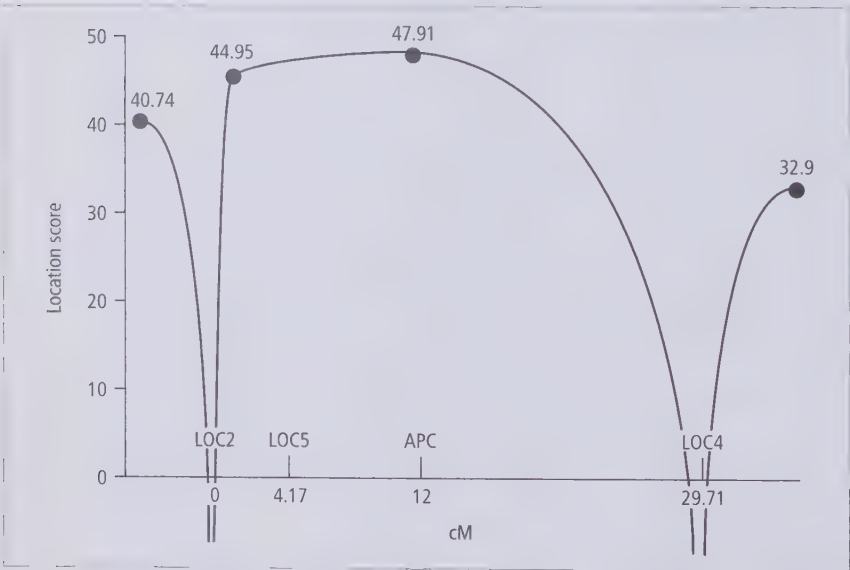


Fig. 3.12 Graphical representation of LINKMAP output. This must usually be drawn by hand. Reproduced with permission from [38].

Protocol 5 Pinpointing a disease gene with LINKMAP

Use IGD/X-PED to generate files for LINKAGE, called `apc_mult.ped` (the pedigree file) and `apc_mult.par` (the parameter file), using Protocol 2 as described in Section 3.3.8. See Fig. 3.4 for a discussion of the structure of these files.

To simplify things, use the three markers (`polym2`, `polym4` and `polym5`) that show significant linkage with APC (following a two-point analysis) as the map template. The first thing to do is to create a GLA object containing the three polymorphisms as well as the adenomatous polyposis coli trait. I suggest you call the GLA `apc_mult`. Remember to add the `Pick_me_to_call` line, save the database and then select the `Pick_me_to_call CALL to_linkage` tag.

The two files created (`apc_mult.ped` and `apc_mult.par`) will be in your `~/Link` directory. Transfer to your LINKAGE window and convert the pedigree file with

```
makeped apc_mult.ped apc_mult.ppd n [return]
```

In the following dialogue, the names of the loci (`polym2`, `polym4` and `polym5`) are referred to by their positions in the linkage parameter file (`p1`, `p2` and `p3`) and the APC locus (the test locus) by `p4`.

Next, at the Unix % prompt, type

```
lcp [return]
```

and edit the first screen to use the files `apc_mult.ppd` as the pedigree file name and `apc_mult.par` as the Parameter file name. Set the Command file name to `apc_mult.sh`. [**ctrl-n**] to the Pedigree Options screen and select General Pedigrees.

Type [**ctrl-n**] to move to the General Pedigree Analysis Options screen and select the LINKMAP option. [**ctrl-n**] again to the LINKMAP—Test Interval Options screen and select All Intervals. [**ctrl-n**] and select the No Sex Difference option, then [**ctrl-n**] to the LINKMAP—Map Specification Command screen. Set the test locus to `p4` and the fixed loci to `p1 p3` and `p2`. Set the recombination fractions to `0.04` and `0.20`. [**ctrl-n**] then [**ctrl-z**] to finish. These values represent the recombination fractions between `p1` and `p3` (0.04) and `p3` and `p2` (0.20).

Execute the script file by typing `apc_mult.sh` [**return**]. This may take some time.

When the job has completed, run LRP with

```
lrp [return].
```

Change the STREAM file name to `apc_mult.stm` at the Input File And Report Title menu and hit [**ctrl-n**]. Select General Pedigree Reports from the Report Options menu and [**ctrl-n**].

Use the cursor keys to select Location Score Report (LINKMAP) from the General Pedigree Report Options menu. [**ctrl-n**] to the Location Score Report (LINKMAP) Formats menu and select Table Format. [**ctrl-n**] to the Report Output Options screen and select Output Report To The Screen. Hit [**ctrl-n**]. LRP takes a few seconds to lay out the report. If the report is more than one page long, [**ctrl-n**] allows you to move a page at a time. [**ctrl-z**] finishes (Fig. 3.11). The results may be represented graphically by a figure (Fig. 3.12).

Troubleshooting

Typing errors and their detection

Much interest has been generated around the issue of genotyping errors in reference databases such as the CEPH. Consortium mapping efforts using data from this database have estimated the residual error rate at somewhat less than 1% [17,18]. This is the estimated error rate after some attempt has been made to identify intralocus recombinants, double recombinants over short distances and families that contain an atypical number of recombinants. The residual error rate puts an upper limit on the resolution of the maps obtainable using this kind of approach. There is some evidence that the data used to construct the maps of Weissenbach and colleagues [24] have a lower error rate than the mixture of RFLP, VNTR and PCR marker systems in the main CEPH database. Some attempt has been made to build the possibility of error within the framework of the likelihood model [50] and Haines [51] has described a method for detecting errors based on the differential inflation in map length by marker systems containing errors. Multiple two-point approaches such as MAP are more robust in the presence of typing errors (N.E. Morton, personal communication).

In the absence of a generally available statistical screening tool for reference data, researchers are advised to be wary of data sets which have not been subject to scrutiny by one of the CEPH consortia. Screening of haplotypes is as important here as anywhere else in linkage analysis, and is greatly facilitated by the ‘chrompic’ option of CRI-MAP (Section 3.3).

It has been stressed that the maps produced using any of the methods described in this chapter should be reconciled with the primary data. Of particular importance is the way in which haplotypes segregate within families. X-PED can help in this process.

Problems with the programs

CRI-MAP core dumps almost always indicate a problem with the CRI-MAP parameter file. CRI-MAP is very sensitive to errors in this file and will not detect or uncover them graciously. If CRI-MAP crashes unexpectedly, look at your parameter file with great care and in particular, look for duplicate loci in the `ordered_loci` field. If CRI-MAP crashes, then your orders database (the `.ord` file) will almost certainly be unreadable and you will need to create a new one from scratch unless you are very confident about the format of the file. Always keep a copy of this file. If necessary, create a new one using the CRI-MAP ‘prepare’ option, described in Section 3.3.

.....

3.4 Integration of physical and genetic maps

An area which is currently receiving a great deal of attention is the integration of genetic and physical maps (see also Chapter 16). Until now, this has mostly been done by mapping the same elements onto maps using other techniques (radiation hybrids, YAC/STS content), taking into account conflicting or supporting evidence on order from these other methods when constructing maps [52]. This technique does not usually embody any rigorous statistical methodology. The location database (ldb) of Morton and colleagues [53] is an attempt to place the integration of maps on a more quantitative, formal basis. ldb is defined in terms of data and algorithms with which diverse mapping evidence can be used to produce a summary map projected on to a mega base scale, based on current estimates of chromosomal length [3]. It has already been used to produce integrated maps of chromosomes 1 [54] and 2 [22] and is available on the Web (http://cedar.genetics.soton.ac.uk/public_html).

Alternative approaches to solving the problem include the System for Integrated Genome Map Assembly (SIGMA) from the Los Alamos National Laboratory (<ftp://atlas.lanl.gov>) in the US and the CPROP algorithm [55]. A more recent development is the characterization of breakpoints in the CEPH families, an approach which shows great promise [56].

Acknowledgements

Some of the material included here was derived from courses sponsored by the UK Human Genome

The techniques described in this chapter are used to:

- manage human genetic data and assist in a linkage analysis
- map new polymorphisms to a small chromosomal region
- construct and improve genetic maps
- use existing genetic maps to localize disease genes

Applications box 3.1

The EUROGEM Project

The European Gene Mapping Project (EUROGEM) was conceived in 1988 by the EC working party on Human Genome Analysis with the objective of improving current linkage maps to a density of 5-cm. The activities included the characterization of new polymorphic markers and extensive genotyping on the CEPH families, followed by the construction of a genome-wide set of genetic maps [16]. The following discussion is largely taken from this paper.

Following a long period of data collection by the EUROGEM laboratories, a process of map construction and error correction was initiated. Each network laboratory was assigned one or more of the 22 autosomal chromosomes or chromosome X. The aim was to place the new EUROGEM markers on a well-supported framework map which could be composed of selected CEPH markers, Généthon CA-repeats or Cooperative Human Linkage Centre (CHLC) systems [25,13].

Laboratories made their own choice as to which framework to choose (CEPH or CHLC) and placed as many as possible of the newly typed EUROGEM markers on the base map with 1000:1 support.

Each laboratory adopted a particular strategy for producing the map, similar to that described in Section 3.3. CRI-MAP version 2.4 [27] was used by all but two of the laboratories, who used MultiMap [46]. Map-building began in parallel with the elimination of allelic exclusions, with corrections being reported so that the data could be corrected and the

CRI-MAP input files regenerated on a continuous basis. Most laboratories used the CRI-MAP ‘build’ routine to enlarge their maps. In the course of map-building, intralocus recombinants were detected and reported back to the contributing laboratories who, in turn, fed corrections back to the central site. Considerable advantage was derived from most sites being connected to the Internet, as communication by electronic mail between labs was mostly very fast, and corrected data files could be downloaded within a few hours of requesting corrections. Additionally, help and advice from those more experienced in map-building could be obtained quickly and easily, and problems with using the software diagnosed and solved. As the maps stabilized, unusual clusters of recombinants could be detected using the ‘chrompic’ option of CRI-MAP, and further rounds of error-checking and correction were carried out. In cases where it was impossible to verify the data and strong doubts still remained as to their validity, they were removed from the data files.

As well as regular checking throughout for local support of at least 1000:1 with CRI-MAP’s ‘flips2’ option, the final maps were checked with ‘flips4’ to ensure that alternative orders had been sufficiently explored. For those markers that could not be placed with 1000:1 support, CRI-MAP ‘all’ runs were conducted for each to determine the range of positions that fell within the 3-unit support interval, and the support intervals for these markers were plotted separately. As a representative example, Fig.3.1 shows the framework map of human chromosome 2, illustrating the order of markers supported at 1000:1 and Fig.3.2 shows the approximate map of the same chromosome [57].

Case Study 3.1

Mapping Project under the auspices of the UK Medical Research Council. ICRF provided support during the preparation of this manuscript.

References

- 1 Edwards, A.W.F. (1972) *Likelihood* (Cambridge University Press, Cambridge).
- 2 Pascoe, L. & Morton, N. (1987) The use of map functions in multipoint mapping. *Am. J. Hum. Genet.* **40**, 174–183.
- 3 Morton, N.E. (1991) Parameters of the human genome. *Proc. Natl. Acad. Sci. USA* **88**, 7474–7476.
- 4 Bouzyk, M., Evans, C., Cullin, L. *et al.* (1996) Isolation and mapping of three new polymorphic markers to chromosomes 3, 20 and 21. *Annls Hum. Genet.* **60**, 73–77.
- 5 Deckert, J., Nothen, M.M., Bryant, S.P. *et al.* (1995) Human adenosine A1 receptor gene: Systematic screening for DNA sequence variation and linkage mapping on chromosome 1q31–32.1 using a silent polymorphism in the coding region. *Biochem. Biophys. Res. Commun.* **214** (2), 614–621.
- 6 Davies, J., Kawaguchi, Y., Bennett, S.T. *et al.* (1994) A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* **371**, 130–136.
- 7 Keats, B.J.B., Sherman, S.L., Morton, N.E. *et al.* (1990) Guidelines for human linkage maps: an international system for human linkage maps (ISLM, 1990). *Genomics* **9**, 557–560.
- 8 Botstein, D., White, R.L., Skolnick, M. & Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314–331.
- 9 Ott, J. (1974) Estimation of the recombination in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am. J. Hum. Genet.* **26**, 588–597.
- 10 Lathrop, G.M. & Lalouel, J.M. (1984) Easy calculations of lod scores and genetic risks on small computers. *Am. J. Hum. Genet.* **36**, 460–465.
- 11 Lander, E.S. & Green, P. (1987) Construction of multilocus genetic linkage maps in humans. *Proc. Natl. Acad. Sci. USA* **84**, 2363–2367.
- 12 Donis-Keller, H., Green, P., Helms, C. *et al.* (1987) A genetic linkage map of the human genome. *Cell* **51**, 319–337.
- 13 Buetow, K.H., Weber, J.L., Lundwigsen, S. *et al.* (1994) Integrated human genome-wide maps constructed using the CEPH reference panel. *Nature Genet.* **6**, 391–393.
- 14 CHLC (1994) A comprehensive human linkage map with centimorgan density. *Science* **265**, 2049–2054.
- 15 NIH/CEPH collaborative mapping group. (1992) A comprehensive genetic linkage map of the human genome. *Science* **258**, 67–68.
- 16 Spurr, N.K., Bryant, S.P., Attwood, J. *et al.* (1994) European Gene Mapping Project (EUROGEM): genetic maps based on the CEPH reference families. *Eur. J. Hum. Genet.* **2**, 193–252.
- 17 Dracopoli, N.C., O'Connell, P., Elsner, T.I. *et al.* (1991) The CEPH consortium linkage map of human chromosome 1. *Genomics* **9** (4), 686–700.
- 18 Spurr, N., Cox, S., Bryant, S.P. *et al.* (1992) The CEPH consortium linkage map of human chromosome 2. *Genomics* **14**, 1055–1063.
- 19 Attwood, J., Chiano, M., Collins, A. *et al.* (1994) The CEPH consortium map of chromosome 9. *Genomics* **19**, 203–214.
- 20 Bowcock, A.M., Gerken, S.C., Barnes, R.I. *et al.* (1993) CEPH consortium linkage map of human chromosome 15q. *Genomics* **14**, 833–840.
- 22 Cox, S., Bryant, S.P., Collins, A. *et al.* (1995) Integrated genetic map of human chromosome 2. *Annls Hum. Genet.* **59**, 413–434.
- 23 Dausset, J., Cann, H., Cohen, D. *et al.* (1990) Centre d'étude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* **6**, 575–577.
- 24 Weissenbach, J., Gyapay, G., Dib, C. *et al.* (1992) A second-generation linkage map of the human genome. *Nature* **359**, 794–801.
- 25 Gyapay, G., Morissette, J., Vignal, A. *et al.* (1994) The 1993–94 Génethon human genetic linkage map. *Nature Genet.* **7**, 246–339.
- 26 Cooperative Human Linkage Center (1993) CHLC Report **1** (1).
- 27 Green, P., Falls, K. & Crooks, S. (1989) Documentation for CRI-MAP, version 2.4. Available from Phil Green: phg@u.washington.edu.
- 28 Thompson, E.A. (1987) Crossover counts and likelihood in multipoint linkage analysis. *IMA J. Math. Appl. Med. Biol.* **4**, 93–108.
- 29 Hultén, M. & Lindsten, J. (1973) Cytogenetic aspects of human male meiosis. *Adv. Hum. Genet.* **4**, 327–387.
- 30 Thompson, E.A. (1986) *Pedigree Analysis in Human Genetics* (The Johns Hopkins University Press, Baltimore).
- 31 Elston, R.C. & Stewart, J. (1971) A general model for the genetic analysis of pedigree data. *Hum. Hered.* **21**, 523–542.
- 32 Lange, K. & Elston, R.C. (1975) Extensions to pedigree analysis I. Likelihood calculation for simple and complex pedigrees. *Hum. Hered.* **25**, 95–105.
- 33 Cannings, C., Thompson, E.A. & Skolnick, M.H. (1978) Probability functions on complex pedigrees. *Adv. Appl. Prob.* **10**, 26–61.
- 34 Morton, N.E. (1955) Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* **7**, 277–318.
- 35 Morton, N.E. & Andrews, V. (1989) MAP, An expert system for multiple pairwise linkage analysis. *Annls Hum. Genet.* **53**, 263–269.
- 36 Curtis, D. & Gurling, H. (1993) A procedure for combining 2-point lod scores into a summary multipoint map. *Hum. Hered.* **43**, 173–185.
- 37 Bryant, S. (1991) Software for genetic linkage analysis. In *Protocols in Human Molecular Genetics. Methods in Molecular Biology* (Mathew, C., Ed.) Vol. **9**, 403–418 (Humana Press, Clifton, NJ).
- 38 Bryant, S. (1994) Genetic Linkage Analysis. In *Guide to Human Genome Computing* (Bishop, M.J.B., Ed.) 59–110 (Academic Press, London).
- 39 Bryant, S. (1996) Software for genetic linkage analysis: an update. *Mol. Biotech.* **5** (1), 49–61.
- 40 Kwiatkowski, D.J., Dib, C., Slaugenhaupt, S.A. *et al.*

- (1993) An index marker map of chromosome 9 provides strong evidence for positive interference. *Am. J. Hum. Genet.* **53**, 1279–1288.
- 41 Lathrop, G.M., Lalouel, J.M., Julier, C. & Ott, J. (1985) Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. *Am. J. Hum. Genet.* **37**, 482–498.
- 42 Ritter, O., Kocab, P., Senger, M., Wolf, D. & Suhai, S. (1994) Prototype implementation of the integrated genomic database. *Comput. Biomed. Res.* **27**, 97–115.
- 43 Dunham, I., Durbin, R., Thierry-Mieg, J. & Bentley, D.R. (1994) Physical mapping projects and AceDB. In *Guide to Human Genome Computing* (Bishop, M.J., Ed.) 111–158 (Academic Press, London).
- 44 Spiridou, A., Spurr, N.K. & Bryant, S. (in preparation) IGD/X-PED: A system for the management and graphical display of pedigree data.
- 45 Spiridou, A., Spurr, N.K. & Bryant, S. (in preparation) IGD: Facilitating genetic linkage analysis.
- 46 Matisse, T.C., Perlin, M. & Chakravarti, A. (1994) Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nature Genet.* **6**, 384–390.
- 47 Terwilliger, J. & Ott, J. (1994) *Handbook of Human Genetic Linkage* (The Johns Hopkins University Press, Baltimore).
- 48 Ott, J. (1991) *Analysis of Human Genetic Linkage* Revised edn (Johns Hopkins University Press, Baltimore).
- 49 Bodmer, W.F., Bailey, C.J., Bodmer, J. *et al.* (1987) Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature* **328**, 614–616.
- 50 Lincoln, S.E. & Lander, E.S. (1992) Systematic detection of errors in genetic linkage data. *Genomics* **14**, 604–610.
- 51 Haines, J.L. (1992) Chromlook: an interactive program for error detection and mapping in reference families. *Genomics* **14**, 517–519.
- 52 Doggett, N.A., Goodwin, L.A., Tesmer, J.G. *et al.* (1995) An integrated physical map of human chromosome 16. *Nature* **377** (Suppl.), 335–365.
- 53 Morton, N.E., Collins, A., Lawrence, S. & Shields, D.C. (1992) Algorithms for a location database. *Annls Hum. Genet.* **56**, 223–232.
- 54 Collins, A., Keats, B.J., Dracopoli, N. *et al.* (1992) Integration of gene maps: chromosome 1. *Proc. Natl. Acad. Sci. USA* **89**, 4598–4602.
- 55 Letovsky, S. & Berlyn, M.B. (1992) CPROP: a rule-based program for constructing genetic maps. *Genomics* **12**, 435–446.
- 56 Fain, P.R., Kort, E.N., Chance, P.F. *et al.* (1995) A 2D-crossover map of the X: Foundation for crossover mapping and the integration of physical and genetic maps.
- 57 Flint, T.J., Hertz, J., Vergnaud, G. *et al.* (1994) The EUROGEN Map of human chromosome 2. *Eur. J. Hum. Genet.* **2**, 206–207.

Chapter 4

Automated map construction: MultiMap

Tara Cox Matise & Aravinda Chakravarti*

Department of Statistical Genetics, The Rockefeller University, 1230 York Avenue, Box 192, New York, NY 10021, USA

*Department of Genetics, Room BRB 721, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA

4.1	Introduction, 90	4.2.3	Documentation, 94
4.2	Obtaining and running MultiMap, 94	4.2.4	Mailing list, 95
4.2.1	What hardware is necessary?, 94		Troubleshooting, 95
4.2.2	What software is necessary and how do I get it?, 94		MultiMap crashes, 95
4.2.2.1	Lisp, 94		References, 95
4.2.2.2	MultiMap and CRI-MAP, 94		

4.1 Introduction

Genetic linkage maps are vital for the positional cloning of genes and for the identification of at-risk individuals within families segregating genetic disorders. The genetic maps necessary for these applications must be accurate at high resolution since a resolution of 1–2cM is necessary for the precise localization of disease genes, particularly those that predispose to complex disorders (see Chapter 2) [1]. Classically, the pace of genetic mapping has been slow since the number of polymorphic markers originally available was limited. The advent of molecular genetics has enabled the purposeful identification, characterization and mapping of thousands of highly polymorphic markers (see Chapter 5) to construct genome-wide linkage maps (see Chapters 3 and 16). The discovery of widespread polymorphisms at microsatellite motifs, easily assayable by the polymerase chain reaction (PCR) and whose characteristics can be electronically disseminated, has led to the construction of whole-genome linkage maps with high marker resolution. This has been made possible in particular by the availability of reference human pedigrees through the Centre d’Etude du Polymorphisme Humain (CEPH) [2] (see Appendix V for address). This panel of 65 three-generation families, DNA samples from which are provided to collaborating investigators, has allowed the genetic mapping of human chromosomes on a common set of families. Figure 4.1 shows the number of genetic markers with heterozygosity greater than 70% available in the Human Genome Data Base (GDB) (see Chapter 37) per year since 1989.

As of January 1996, there were 11 616 polymorphic

PCR-based markers reported in GDB, 4720 of which have heterozygosity greater than 70%. Accordingly, the number of high-resolution maps is also increasing, as evidenced by several recent publications of genome maps of humans [3,4], mice [5], cows [8], and other organisms (see the chapters in Section V).

As the number of markers available for mapping increases, so does the complexity of multipoint map construction. Several computer programs are currently available for computation of the genetic likelihood of specified ordered sets of markers, including LINKAGE [7], MAPMAKER [8], CRI-MAP [9], MENDEL [10], and FASTLINK [11,12]. The only way to determine the most likely order of a set of n markers is to compute and compare the likelihoods of all $n!/2$ possible locus orders. However, computation of likelihoods is so time-consuming that this is not feasible unless n is very small. Therefore, genetic mapping relies on heuristics to determine a set of candidate orders which have a high probability of containing the most likely order.

With the exception of CRI-MAP’s ‘build’ routine, the available programs for likelihood computation do not include algorithms for map construction but are algorithms for computing the likelihood of a specified locus order. In other words, the user must determine an order of markers to be analysed, run one of the programs to compute the likelihood of that order, choose another order of markers, compute its likelihood, and compare the two likelihoods to determine the more likely order. This process is repeated with overlapping sets of linked markers until a complete map has been constructed. With the number of markers currently available, map construction for one chromosome can easily require hundreds of user-directed steps. Such a process is time-consuming, tedious and error-prone.

Because the procedure for linkage map construction can follow very specific rules, it is particularly amenable to automation of the computational steps. We have written the computer program MultiMap to enable the automated construction of genetic linkage maps. MultiMap implements a particular set of mapping algorithms developed, tested and used in our mapping projects. The program incorporates one algorithm for construction of a framework map (Fig. 4.2), when markers are mapped at low resolution and an approximately equal spacing, and a different algorithm for expansion of a framework map into a comprehensive map (Fig. 4.3), when additional markers are added to the map regardless of spacing, in order to increase map resolution. One advantage of MultiMap over CRI-MAP’s ‘build’

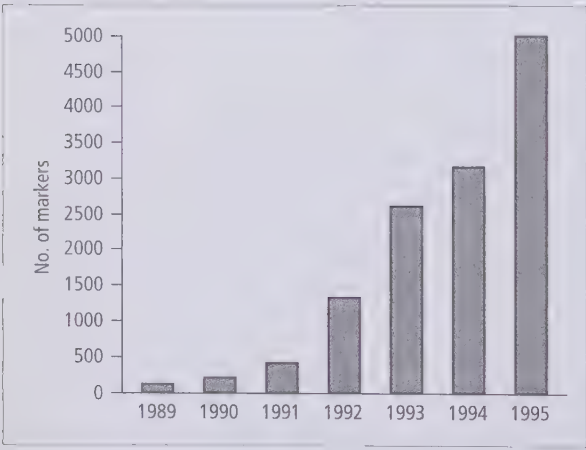


Fig. 4.1 Number of highly polymorphic ($H > 0.7$) markers available in GDB per year since 1989.

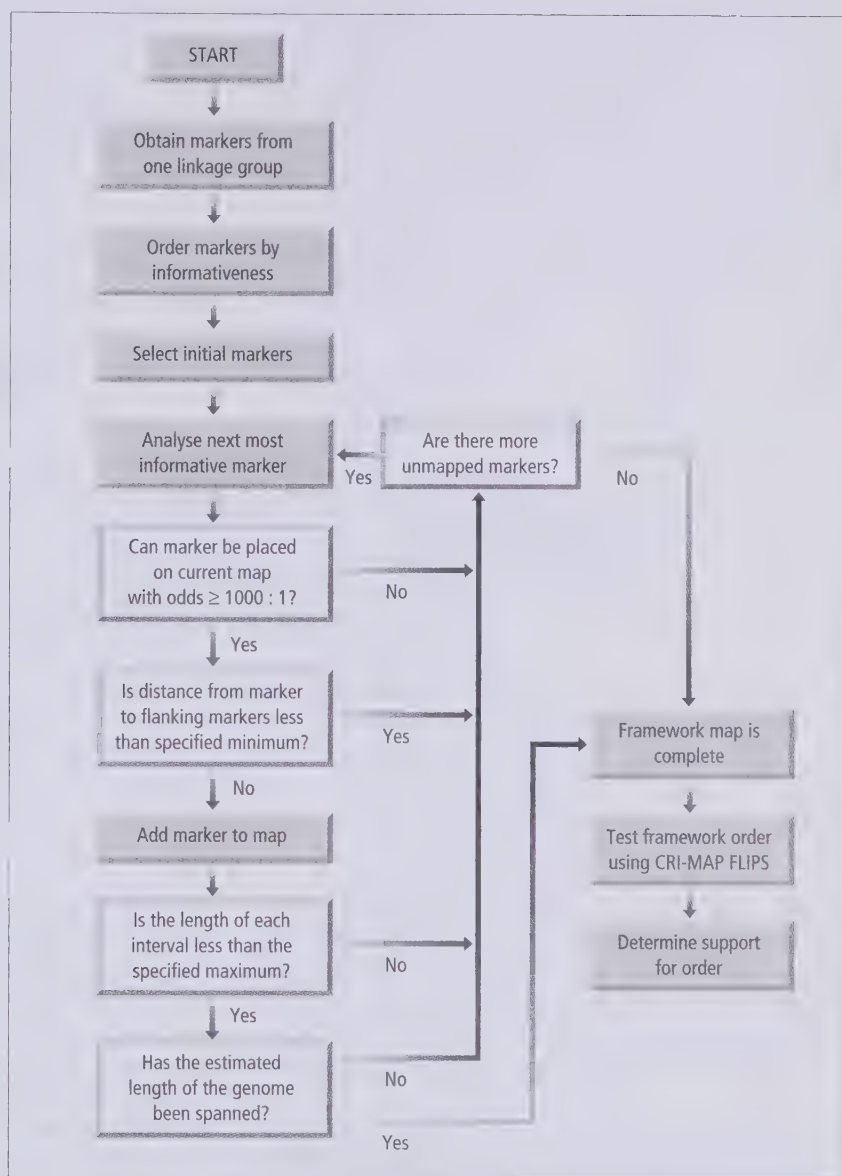


Fig. 4.2 Flowchart of algorithm for framework map construction.

routine is that during the map-building process, MultiMap frequently uses a test to confirm that the markers in the current map are, in fact, in the most likely map order.

In addition, markers are mapped in order of locus content. The content of a marker locus is defined by criteria such as informativeness (heterozygosity or pairwise joint-polymorphism information content (PIC) value [13]), relative ease of genotyping (i.e. Southern blotting vs. PCR-based markers), marker quality score (based on background, ease of allele identification, etc.), ability to be multiplexed with other markers, and map distance between a marker and its nearest neighbours. On the basis of these criteria, markers can enter the map-building procedure in a non-random manner, with those markers

with the most desirable characteristics added to the map before others. Currently, MultiMap uses only informativeness and map distance to determine locus content. The ideal linkage map consists of markers with the most desirable locus content; as new markers with better characteristics are developed maps can be reconstructed. Details of these algorithms are given in Matise *et al.* [14] and in the MultiMap documentation (see below).

While any of several programs could have been chosen for computation of multipoint likelihoods, MultiMap uses CRI-MAP since it is computationally faster, can analyse larger data sets than many of the other available programs, and is easily ported to different computing platforms. CRI-MAP is a C program package for likelihood calculations in

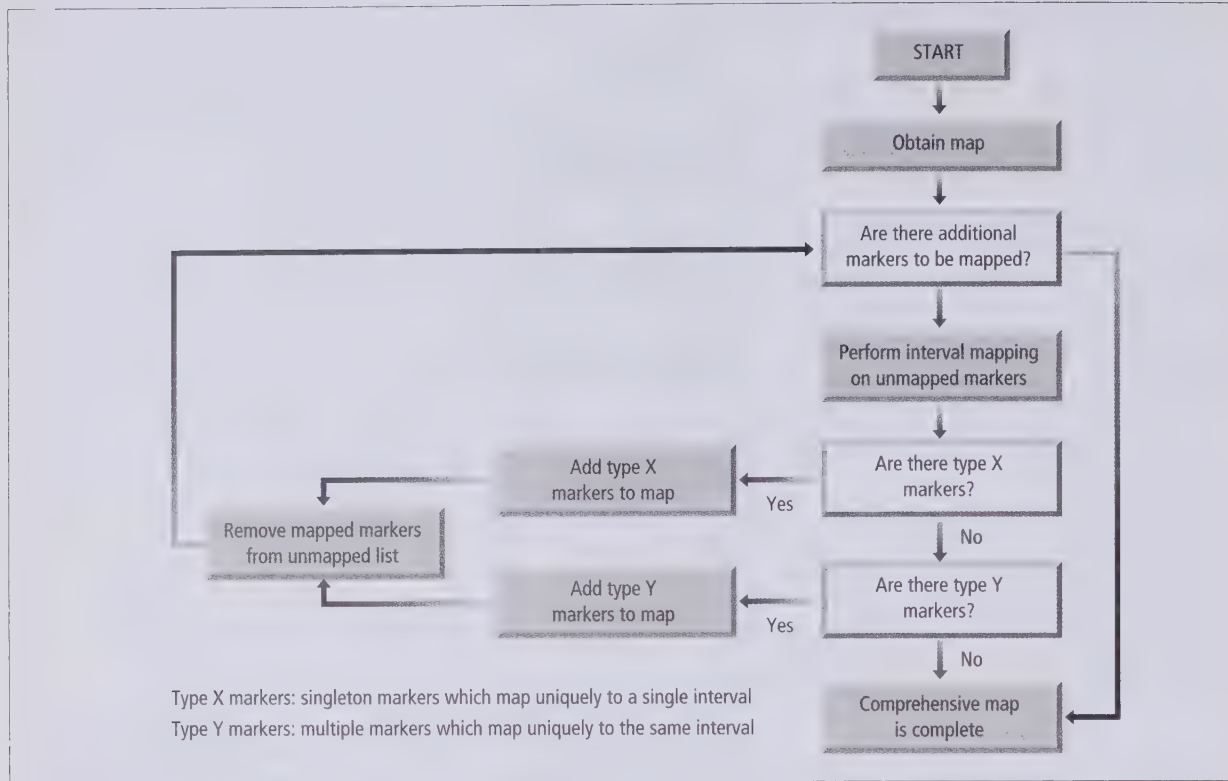


Fig. 4.3 Flowchart of algorithm for comprehensive map construction.

general pedigrees without inbreeding, for any number of codominant loci, and can assume sex difference in recombination. It uses a rapid, layered expectation maximization (EM) algorithm and switch algebras for likelihood calculations, and is very efficient for map construction of large numbers

of marker loci with large numbers of alleles, particularly for CEPH-structured pedigrees (see Chapter 2). CRI-MAP does ignore some linkage information in pedigrees of arbitrary structure, specifically in the presence of incomplete marker genotypes; the LINKAGE [7] program is more relevant to these situations. Nevertheless, even in general pedigrees, the loss of information is estimated to be only 4% for multilocus analysis [15].

MultiMap can aid in several analyses beyond the construction of linkage maps. These additional automated analyses are geared toward description of the markers and the map, further validation of the map, assessment of its accuracy, and analyses of biological features of the map. Some of these features include computation of all pairwise two-point likelihoods and recombination fractions, estimation of marker allele frequencies, estimation of locus heterozygosity, PIC value [16] and pairwise joint-PIC value [13], estimation of chromosome length given a set of unmapped chromosome-specific markers [17], determination of linkage groups, computation of all pairwise lod scores, computation of local support for order, estimation of undetected genotype error rate [18,19], identification of marker loci suspected to contain errors, analysis of sex difference in recombination and its

MultiMap is used to:

- construct genetic linkage maps of codominant DNA markers using automated algorithms by
- constructing framework maps, or
- adding markers to a framework map to construct comprehensive maps
- estimate marker allele frequencies
- estimate marker informativeness by heterozygosity and PIC value
- estimate chromosome length
- identify linkage groups
- compute all pairwise lod scores
- compute local support for order of a marker order
- estimate undetected genotype error rate
- identify marker loci suspected to contain genotyping errors
- analyse parental origin differences in recombination
- for radiation hybrid mapping (see Chapter 14)

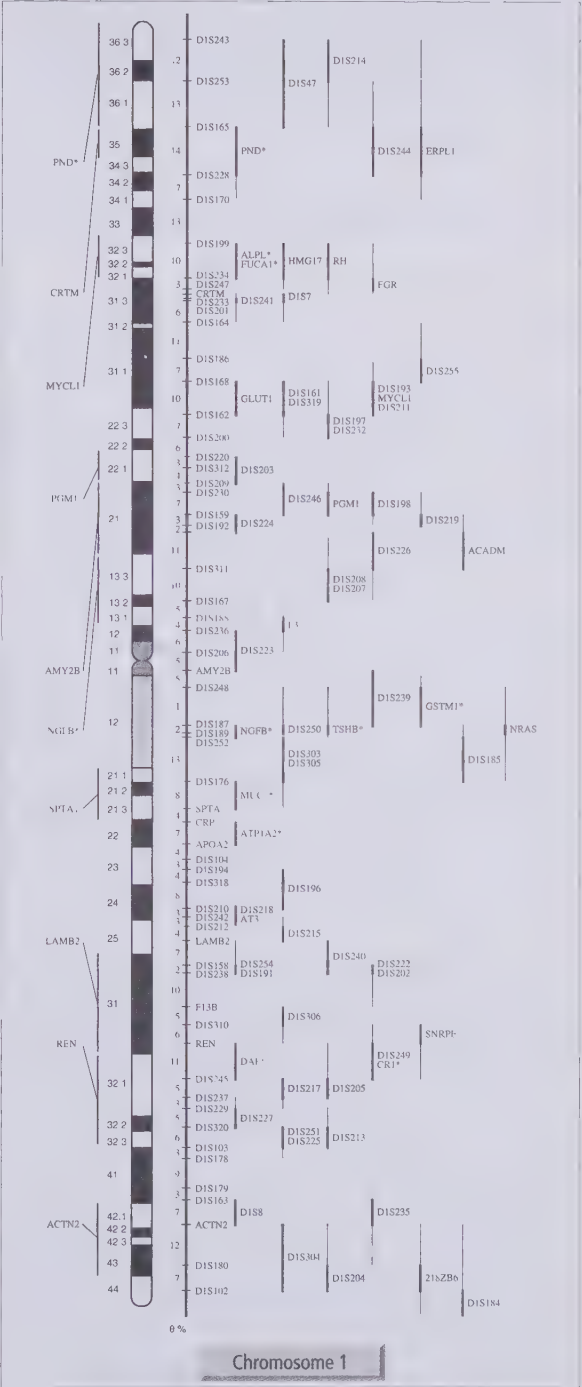
variation along the chromosome, and, analysis of chiasma interference.

There are many advantages to be gained through automation, as implemented in the MultiMap expert system. With the ongoing search for disease genes, particularly those that cause common complex disorders, improvements in the area of high-resolution map construction are needed. In addition to practical issues such as increased speed and greater accuracy, the ability to efficiently reconstruct maps, based on locus content, will become more important as the number and usefulness of genetic markers increase. Because of the increasing awareness and concerns about genotyping errors, constructing maps of these new markers *de novo* will reduce the propagation of errors and will further aid in the detection of errors in older marker genotypes. Similarly, consensus linkage maps constructed by a collaboration of investigators, such as chromosome committees or consortia, can now be more easily, rapidly and accurately constructed *de novo* using the basic genotype data and MultiMap, rather than by manual integration of maps.

MultiMap is suitable for a variety of projects. Its primary use has been in the construction of whole-genome or chromosome maps based on genotype data on a very large number of markers that arise from a few sources (see, for example, Fig. 4.4). MultiMap has also been adapted so that it can be applied to radiation hybrid mapping [20] (see Chapter 14). However, we have found it extremely useful as a tool for synthesizing all available marker data in a single genomic region into a consensus map. It may require some effort to obtain and install a Lisp interpreter (see below), but the time and effort saved over manual methods is considerable, particularly for map construction using default parameter values. To achieve even greater speed during large-scale mapping projects, we have recently implemented a version of CRI-MAP that computes likelihoods in parallel using a distributed network

of workstations [21] in conjunction with parallel virtual machine (PVM) software [22]. With additional software installation, MultiMap can interface with the parallel version of CRI-MAP. MultiMap may not be most suitable for small or infrequent mapping projects. In these cases, one could simply apply one's own mapping algorithm using one of the available programs for likelihood computation, such as CRI-MAP.

Fig. 4.4 Example of an annotated sex-averaged human genetic linkage map of chromosome 1 produced using MultiMap (from ref. 14). More up to date maps have now been published from other sources [4,5]. All loci are localized to positions with 1000:1 odds or greater, with haplotyped markers denoted by an asterisk (*). The loci on the thick line are uniquely localized; the interlocus distances provided are percentages of sex-equal maximum likelihood recombination values. Numerical values of 1 cM or less are not shown. The loci to the right are all shown in their 1000:1 odds positions which span more than one primary interval; the thick bar shows the most likely primary map interval. From ref. 14, with permission.



The assembly of a consensus map for chromosome 1

One useful application of MultiMap is as a tool for assembling a map that is the consensus of several maps. For example, prior to a recent human chromosome 1 workshop [24], over 200 markers had been mapped to this chromosome by several groups, including Engelstein *et al.* [25], CHLC [3], Généthron [26] and the NIH/CEPH Collaborative Mapping Group [27]. Since none of these maps contained the entire set of markers, it was quite difficult to determine the most likely order or map distances for any arbitrary set of chromosome 1 markers. Therefore, for the chromosome 1 workshop, one map of chromosome 1 was constructed *de novo*, using MultiMap, which contained the entire set of markers covered by the previously published maps.

Case Study 4.1**4.2 Obtaining and running MultiMap****4.2.1 What hardware is necessary?**

MultiMap can be run on any computer for which both a Lisp interpreter and a standard C compiler are available. C compilers are available for most computers, but Lisp interpreters are less widely portable. MultiMap has been used on specific Sun, HP and Dec workstations. Section 4.2.2.1 below provides further details.

4.2.2 What software is necessary and how do I get it?

You must install a Lisp interpreter, MultiMap, and an edited version of CRI-MAP version 2.41. If you wish to use the parallel version of CRI-MAP with MultiMap, you must obtain the CRI-MAP-PVM software, which is distributed at the same file transfer protocol (ftp) site as MultiMap. Additional software is required for radiation hybrid mapping.

4.2.2.1 Lisp

There are two software components to MultiMap: the MultiMap code, which implements our map-building algorithm, and CRI-MAP [9], which is used by MultiMap for likelihood calculations. CRI-MAP is written in the C computer language, and can be compiled and run on any machine with a standard C compiler. MultiMap is written in the computer language Common Lisp. Lisp programs are developed using an interpreter, which is similar to a compiler, except that it provides an interactive working environment with its own shell. Unlike languages like C, Pascal and Fortran, it is not usually possible to run Lisp code in the absence of a Lisp interpreter. Therefore, MultiMap can only be

run on machines on which a Lisp interpreter is installed.

There are several commercial and non-commercial Lisp interpreters available, and MultiMap is available compiled under one of each type. We recommend the use of CLISP [23], which is available at the anonymous ftp site [ma2s2.mathematik.uni.karlsruhe.de, in the directory pub/lisp/clisp/binaries](http://ma2s2.mathematik.uni.karlsruhe.de/pub/lisp/clisp/binaries). It can be run on several different machine architectures, and MultiMap has been tested using CLISP on the following platforms: sun4-sunos51, sun4-sunos4, hp9000s800, dec5000-ultrix, and dec-Alpha.

4.2.2.2 MultiMap and CRI-MAP

MultiMap and CRI-MAP are available electronically via the Internet. Either the ftp or the World Wide Web (WWW) can be used to retrieve these programs (see below). Once you have connected to our distribution site, you should first retrieve and read the readme file, which provides all necessary further instructions. If you are unable to retrieve the program electronically, please contact us to pursue other options.

Note that this version of CRI-MAP (version 2.41) has been edited from the original version of CRI-MAP for the purpose of integration of MultiMap only. The algorithms for computation of genetic likelihoods have not been changed, but this version is not useable outside MultiMap. To obtain the original CRI-MAP version, contact Phil Green via e-mail at phg@u.washington.edu.

Access via anonymous ftp (see Chapter 35) The MultiMap ftp address is linkage.rockefeller.edu. Enter anonymous as your logname, and your e-mail address as your password. Use the 'cd' command to change to the multimap directory [`cd multimap`]. The ftp site is mirrored in Cambridge, UK at the European Bioinformatics Institute (EBI). Their address is ftp.ebi.ac.uk, and the directory is `/pub/software/linkage_and_mapping/MULTIMAP`.

Access via the World Wide Web (see Chapter 35) The address for the MultiMap home page is <http://linkage.rockefeller.edu/multimap>.

4.2.3 Documentation

Extensive documentation has been written for the MultiMap program. It is obtainable in the same manner as the MultiMap and CRI-MAP code (see Section 4.2.2.2). It is available in four formats: Microsoft Word (Macintosh), PostScript, text, and in HTML format for the WWW.

4.2.4 Mailing list

We keep a mailing list with e-mail addresses of all users. This is our only means of communicating bug reports and updates. We consider it a vital link to our

users and ask each user to e-mail a request to us asking to be added to this list. We do not have any other means of identifying users; simply retrieving the program does not automatically add you to the mailing list.

Troubleshooting

MultiMap crashes

As with any computer program, there are many situations which might cause MultiMap not to run. Many of these are explicitly addressed in the MultiMap documentation. Others are best addressed when they occur by sending a detailed report to: multimap@chimera.gene.cwru.edu.

References

1 Lander, E.S. & Schork, N.J. (1994) Genetic dissection of complex traits. *Science* **265**, 2037–48.

2 Dausset, J. (1986) Le centre d'étude du polymorphisme humain. *La Presse Médicale* **15**, 1801–1802.

3 Cooperative Human Linkage Center (CHLC), Murray, J.C., Buetow, K.H., Weber, J.L. *et al.* (1994) A comprehensive human linkage map with centimorgan density. *Science* **265**, 2049–2054.

4 Dib, C., Fauré, S., Fizames, C. *et al.* (1996) A comprehensive genetic map of the human genome based on 5264 microsatellites. *Nature* **380**, 152–154.

5 Dietrich, W.F., Miller, J., Steen, R. *et al.* (1996) A comprehensive map of the mouse genome. *Nature* **380**, 149–152.

6 Barendse, W., Armitage, S.M., Kossarek, L.M. *et al.* (1994) A genetic linkage map of the bovine genome. *Nature Genet.* **6**, 227–235.

7 Lathrop, G.M., Lalouel, J.M., Julier, C & Ott, J. (1984) Strategies for multilocus linkage analysis in humans. *Proc. Natl Acad. Sci. USA* **81**, 3443–3446.

8 Lander, E.S., Green, P., Abrahamson, J. *et al.* (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174–181.

9 Lander, E.S. & Green, P. (1987) Construction of multi-locus genetic linkage maps in humans. *Proc. Natl Acad. Sci. USA* **84**, 2363–2367.

10 Lange, K., Weeks, D. & Boehnke, M. (1988) Programs for pedigree analysis: MENDEL, FISHER and dGENE. *Genet. Epidemiol.* **5**, 471–472.

11 Cottingham, R.W., Idury, R.M. & Schaeffer, A.A. (1993) Faster sequential genetic linkage computations. *Am. J. Hum. Genet.* **53**, 252–263.

12 Schäffer, A.A., Gupta, S.K., Shiram, K. & Cottingham, R.W. (1994) Avoiding recomputation in linkage analysis. *Hum. Hered.* **44**, 225–237.

13 Chakravarti, A. (1991) Information content of the Centre d'Etude du Polymorphisme Humain (CEPH) family structures for linkage studies. *Hum. Genet.* **87**, 721–724.

14 Matisse, T., Perlin, M. & Chakravarti, A. (1994) Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nature Genet.* **6**, 384–390.

15 Lupski, J.R., Montes de Oca-Luna, R., Slangenaupt, S.A. *et al.* (1991) DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell* **66**, 219–232.

16 Botstein, D., White, R.L., Skolnick, M. & Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314–331.

17 Chakravarti, A., Lasher, L.K. & Reefer, J.E. (1991) A maximum likelihood method for estimating genome length using linkage data. *Genetics* **128**, 175–182.

18 Buetow, K.H. (1991) Influence of aberrant observations on high-resolution linkage analysis outcomes. *Am. J. Hum. Genet.* **49**, 985–994.

19 Lasher, L., Reefer, J. & Chakravarti, A. (1991) Effects of genotyping errors on the estimation of chromosome map length. *Am. J. Hum. Genet.* **49** (Suppl.), 369.

20 Matisse, T.C. & Chakravarti, A. (1995) Automated construction of radiation hybrid maps using MultiMap. *Am. J. Hum. Genet.* **57** (Suppl.): A15 (Abstract).

21 Matisse, T.C., Schroeder, M.D., Chiavulli, D.M. & Weeks, D.E. (1995) Parallel computation of genetic likelihoods using CRI-Map, PVM, and a network of distributed workstations. *Hum. Hered.* **45**, 103–116.

-
- 22 Geist, G.A. *et al.* (1993) *PVM 3 User's Guide and Reference Manual* (Oak Ridge National Laboratory: Oak Ridge, Tennessee).
- 23 Haible, B. & Stoll, M. (1993) CLISP. © 1992, 1993, 1994 Bruno Haible & Michael Stoll. Karlsruhe, Germany.
- 24 Dracopoli, N.C., Bruns, G.A., Brodeur, G.M. *et al.* (1994) Report & Abstracts of the First International Workshop on Human Chromosome 1 Mapping, 1994. *Cytogenet. Cell Genet.* **67**, 144–165.
- 25 Engelstein, M., Hudson, T.J., Lane, J.M. *et al.* (1993) A PCR based linkage map of human chromosome 1. *Genomics* **15**, 251–258.
- 26 Weissenbach, J., Gyapay, G., Dib, C. *et al.* (1992) A second-generation linkage map of the human genome. *Nature* **359**, 794–801.
- 27 NIH/CEPH Collaborative Mapping Group (1992) A comprehensive genetic linkage map of the human genome. *Science* **258**, 67–86.

Chapter 5

Identification and
characterization of DNA
polymorphisms

John A.L. Armour

*Department of Genetics, School of Medicine, Queen's Medical Centre,
Nottingham NG7 2UH, UK*

5.1	Introduction, 99	
5.1.1	The need for polymorphism: horses for courses, 99	
5.1.2	General properties of polymorphisms, 99	
5.1.2.1	Informativeness, 99	
5.1.2.2	Map placement, 100	
5.1.2.3	Special cases, 100	
5.2	Finding polymorphisms: the easy way, 100	
5.2.1	Map placement, 100	
5.2.1.1	Génethon maps, 101	
5.2.1.2	Other maps, 101	
5.2.1.3	Earlier maps, 101	
5.2.2	More maps to come, 102	
5.2.3	From published map to useful polymorphism, 102	
5.3	Methods for identifying new polymorphisms, 102	
5.3.1	Starting with a defined clone or sequence, 102	
5.3.1.1	Identification of RFLPs by Southern blot hybridization, 103	
5.3.1.2	Finding substitutional polymorphism: SSCP analysis, 104	
5.3.1.3	Sequence analysis, 104	
5.3.1.4	Conversion of sequence polymorphisms to convenient assays, 106	
5.3.2	Starting with large DNA collections, 107	
5.3.2.1	Starting points for subgenomic libraries, 107	
5.3.2.2	Tandem repeat variability: VNTRs, 108	
5.3.2.3	Screening for polymorphism: general considerations, 110	
5.3.2.4	Mononucleotide repeats, 110	
5.3.2.5	Dinucleotide repeat loci, 110	
5.3.2.6	Tri- and tetranucleotide repeats, 111	
5.3.2.7	Minisatellites, 112	
5.4	Multilocus methodology, 113	
5.4.1	DNA fingerprinting: general principles, 113	
5.4.2	Random amplified polymorphic DNA PCR, 113	
5.4.3	Genomic mismatch scanning, 113	
5.4.4	Representational difference analysis, 113	
5.4.5	AFLP scanning, 113	
	Protocol 6	
	Pre-association of dispersed repeat elements, 114	
	Troubleshooting, 115	
	No signal visible, 115	
	Smeared signal, 115	
	Protocol 7	
	SSCP analysis for polymorphism, 115	
	Troubleshooting, 117	
	No bands seen, 117	
	Too many bands, 117	
	Protocol 8	
	Typing amplified RFLPs (PCR-RFLPs), 118	
	Troubleshooting, 118	
	No fragments seen, 118	

Too many fragments, 118
Incomplete digestion, 119

Protocol 9

Typing dinucleotide repeat loci, 119

Troubleshooting, 120

No signal/faint signal, 120

Too many bands: band 'spreading', 121
Too many bands: additional 'constant' bands, 121

Acknowledgements, 121

References, 121

5.1 Introduction

5.1.1 The need for polymorphism: horses for courses

Genetic analysis depends absolutely on the existence of genetic polymorphism, only a minute fraction of which is reflected in visible phenotypic differences. Among human genetic polymorphisms, polymorphisms of DNA sequence are by far the most abundant and convenient for most purposes (Table 5.1). There are a number of uses to which polymorphic loci can be put (see Applications box 5.1), and the precise nature of each project will determine which polymorphisms are most appropriate to the application in hand.

For example, linkage analysis in pedigrees (Chapter 1) simply requires a polymorphic system in which different genotypes can be scored unambiguously. In contrast, the study of loss of heterozygosity from tumour DNA [1] requires polymorphisms that can be typed directly on genomic DNA without distorting the relative dosage of each allele; for this purpose polymorphisms that can be typed by Southern blot hybridization are most simply applicable. Other studies, such as analysis of allelic association, linkage disequilibrium or comparisons of allele frequencies between different populations, require the unambiguous ascertainment of allele identity between

unrelated individuals. This rules out Southern blot typing of some highly polymorphic VNTR loci, at which alleles which are in fact of different origin may appear indistinguishable by length [2].

Some of these cautions only apply in certain contexts, but illustrate the need to match the polymorphic system analysed to the use to which it is to be put. In many cases, the polymorphisms used will mainly be dictated by the genomic location under study, and the nature, availability and practical convenience of established polymorphic loci from that region.

5.1.2 General properties of polymorphisms

For many studies, enough established polymorphic markers will be available from the region under analysis, but other studies will require the identification of new polymorphisms. The practical aspects of these two approaches—‘off the peg’ and ‘DIY’—will occupy most of this chapter. First, however, it is worth briefly considering some general aspects of locus ‘quality’.

5.1.2.1 Informativeness

The most important single property of a polymorphism, in terms of its practical utility, is its informativeness. This generally refers to the frequency with which the two alleles at the locus in any individual can usefully be distinguished; the ability to distinguish the maternally and paternally inherited alleles at a locus lies at the heart of most genetic analysis, including segregation analysis (see Chapter 1). The simplest measure of informativeness is the *heterozygosity*—the frequency in the population of heterozygotes at that locus, which is usually expressed as a percentage or a frequency value between 0 and 1.0. Heterozygosity is a directly predictive measure of usefulness in loss of heterozygosity (LOH) studies in tumours: for example, a locus with heterozygote frequency of 0.6 should allow the assessment of LOH in approximately 60% of tumours.

DNA polymorphisms are commonly used:

- in linkage analysis and to construct reference linkage maps
- in association studies
- to detect loss of heterozygosity (LOH) in tumours

More specialized applications include:

- sib-pair analysis
- confirming identity, twin zygosity or family relationships (see Chapter 6)

Applications box 5.1

Table 5.1 The major classes of DNA polymorphism, with methods of identification and genotyping.

Polymorphism type	Discovery	Typing method(s)
Substitutional	RFLP screening analysis, SSCP, DGGE, etc.	Southern blot, PCR-RFLP, ASO, etc.
Length polymorphism (VNTR)	Hybridization Hybridization	Southern blot (minisatellites, satellites) PCR typing (dinucleotides, STRs)

ASO, allele-specific oligonucleotide; DGGE, denaturing-gradient gel electrophoresis; SSCP, single-stranded conformational polymorphism; STRs, simple tandem repeat array.

In linkage analysis, however, heterozygosity is not such a direct predictor of practical informativeness, and another parameter, the polymorphism information content (PIC), which also varies between 0 and 1.0, incorporates additional information. The problem in segregation analysis is the ambiguity introduced by genotype sharing between heterozygous parents. In a family in which both parents are heterozygous with genotypes 1,2 and 1,2 (Fig. 5.1a), those children of genotype 1,2 (50%, on average) give no information about segregation. More information about segregation in the children is obtained if the parents have genotypes 1,2 and 3,4 (Fig. 5.1b). In the first family it is not possible to tell whether a child with genotype 1,2 (* in Fig. 5.1a) has inherited allele 1 paternally and allele 2 maternally, or vice versa. In the second family the parental origin of the alleles is clear. The expected frequency of uninformative children (as in Fig. 5.1a) is subtracted from the heterozygosity to give the PIC, a better measure of the true informativeness of a polymorphism in linkage analysis [3].

5.1.2.2 Map placement (see also Section 5.2.1)

The results of linkage analysis or other studies involving a polymorphism can be most easily related to a genomic location, and to results obtained with other loci, if the polymorphism has already been localized within the genome. For genetic studies, the most convenient sources of such information are published genetic linkage maps (see below), and such maps can be used to choose appropriate polymorphic loci. The usefulness of a comprehensive map emphasizes the need for other loci, and particularly newly identified polymorphisms, to be placed relative to markers on established maps.

5.1.2.3 Special cases

In some cases the properties of the particular location under analysis give rise to special considerations. Linkage analysis of a genetic disorder located in a subtelomeric region, for example, benefits from

the high concentration of highly informative mini-satellite (VNTR) loci in such regions, and from the high rates of recombination per unit of physical distance, which will lead to the rapid delineation of a relatively short physical interval for the disorder [4,5]. The other side of the coin is that it will be proportionately harder to establish linkage at first using randomly chosen markers. The opposite considerations apply to loci near some centromeres, or other regions where there is relatively infrequent recombination per unit of physical distance: while the relatively short genetic distances make the linkage easier to establish initially, the scarcity of recombinations makes the interval harder to narrow down in subsequent studies—as was the case for Friedrich’s ataxia [6].

Loci on the sex chromosomes show distinctive patterns of inheritance. The X-specific part of the X chromosome will harbour genes for sex-linked disorders; loci in the ‘pseudoautosomal’ X–Y pairing regions [7–9] will be inherited in an apparently autosomal pattern, and although these regions are physically small, they should be considered as possible locations for apparently autosomal genes. Mitochondrial and Y-linked loci are inherited uniparentally, and thus linkage analysis of the genetic disorders associated with these regions is not possible. The involvement of these regions in a phenotype should be evident from the pattern of inheritance, and the sequences responsible for the disorder must be inferred from direct correlations between the phenotype and structural variants in genomic DNA.

5.2 Finding polymorphisms: the easy way

For all but a very few, small regions of the human linkage map large numbers of convenient, highly informative markers have been described. It therefore makes most sense to use these ‘ready-made’ polymorphisms wherever possible. It is only when the selection of ready-made markers is limiting, or seems likely to become so, that it is worthwhile producing them from scratch.

5.2.1 Map placement

Most studies seek to relate linkage (or other) data to existing genetic maps, and thus most sense can be made of results from markers that have already been accurately placed on established genetic maps. This chapter is directed mainly to the study of the human genome, but the general principle—that linkage mapping can proceed most efficiently relative to a

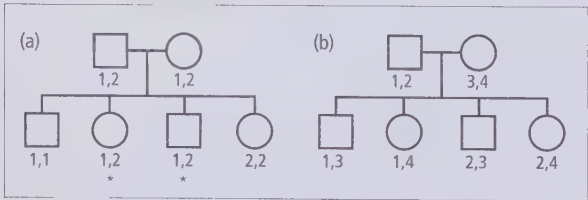


Fig. 5.1 Segregation of alleles from parents into offspring in a simple pedigree. (a) Two alleles; (b) four alleles. For two of the children (*), the parental origin of each allele cannot be traced unambiguously.

framework genetic map—applies in the study of any organism (see Section V). There are a number of sets of human genetic maps currently available, but as a primary reference resource the Généthon maps are unrivalled. Many of the maps described in Section 5.2.1.3 have largely been superseded by more recent productions. They are nevertheless included for the sake of completeness, and since for particular applications and regions the markers used may still have a place. The second Généthon maps [10] have the considerable advantage (particularly to those with limited network access) that the primer sequences and locus characteristics for all the markers used are printed alongside the maps. Note that for genome-wide searches, a subset of microsatellites has been developed for high-throughput linkage analysis using multiplex fluorescent typing in association with genotype analysis software [11].

5.2.1.1 *Généthon maps*

The maps produced by Généthon [10,12,13] are derived from the genotyping of a large number of dinucleotide repeat loci. The development of these markers has in itself made a major contribution to human genome mapping. Particular care has been taken to avoid genotyping errors in generating the data for linkage analysis, and the resulting maps have a high density of coverage, with very few large gaps. The publications for the first two sets of maps [10,12] include printed maps as part of the publication; the most recent maps, comprising an amazing total of 5264 markers, are available electronically [13]. The primer sequences for the loci have also been deposited in GDB (Genome Data Base, see Chapter 37) and (for the second maps) are listed in ref. 10.

The status of the Généthon maps as the unchallenged first point of reference for those mapping human loci has been further bolstered by the appearance of Généthon markers (or corresponding Centre d'Etude du Polymorphisme Humain (CEPH) YACs) in a number of detailed physical maps of the human genome, which thus give independent confirmation of marker order. These include maps based on fluorescent *in situ* hybridization (FISH) [14], radiation hybrid analysis [15], and STS (sequence-tagged site)-content analysis [16] or contig mapping [17] of yeast artificial chromosomes (YACs). A minor potential 'downstream' difficulty in using these maps as a starting point for more detailed linkage analysis is that the placement of Généthon markers relative to markers used in other maps (e.g. those in Section 5.2.1.2) may be unknown. The Généthon maps are

nevertheless generally adopted as the best single starting-point for human linkage analysis.

5.2.1.2 *Other maps*

The Cooperative Human Linkage Center (CHLC) maps [18] include many newly identified markers, with particular emphasis on tri- and tetranucleotide simple tandem repeats. The maps, segregation data and marker descriptions have all been made publicly available by ftp (see below, and Chapter 35). Similarly, more recent maps from Salt Lake City include many useful, newly characterized simple repeat loci [19]. Two publications have reported the generation of integrated maps derived from different data sets [20,21]. The maps produced by Murray *et al.* include data from previous submissions to the CEPH collaboration, as well as new data from large mapping groups, including Généthon, CHLC and the University of Utah. They have the disadvantage for the informationally challenged that the publication does not actually show the maps. However, details of access to ftp sites containing information on the markers, data and maps are given in the report.

Maps deriving from the CEPH collaborations have been published (in this order) for chromosomes 10, 1, 15q, 2, 13, 9, 16, 14 and 11 [22–30]. Their chief benefit is the integration of large numbers of polymorphisms, many of which have appeared previously in other maps, into a single map. The linkage maps usually derive from analysis by many groups of the same data set, and provide an exceptionally well supported framework map for that chromosome. Maps resulting from the collaborative European genetic mapping project (EUROGEM) [31] contain some of the markers analysed by Généthon, and thus these common points of contact can be used to identify specific regions in both maps. They therefore have particular uses in cross-referencing between Généthon markers and other maps.

The assembly of detailed genetic maps from segregation data using breakpoint analysis holds the promise of maps for which detailed marker order can be established with high confidence. This general principle has been illustrated by a detailed breakpoint map of the human X chromosome [32]. Mapping by analysis of individual cross-overs makes more efficient use of the data than multipoint algorithms and allows the identification of presumed data errors; it is likely that this approach will be extended to other chromosomes [33].

5.2.1.3 *Earlier maps*

The Collaborative Research (CRI) maps were the first complete set of published genetic maps to cover

most of the human genome [34]. Their main disadvantages are that in some places (for example, chromosomes 9p, 10p, 14 and 19) the coverage is thin, and some of the markers are relatively uninformative. The markers are restriction fragment length polymorphism (RFLP)/variable number of tandem repeats (VNTR) loci assayed by Southern blot hybridization of genomic DNA using cloned probes, and are thus not as easily 'portable' as PCR-based loci. They nevertheless constitute an important set of polymorphisms, particularly as many of these loci have been integrated into newer maps (see below). For probe requests, information can be obtained from: Bio-medical Products Division, Collaborative Research Incorporated, 2 Oak Park, Bedford, Massachusetts 01730, USA; Tel.: (+1 617) 275-0004, Fax (+1 617) 275-0043.

Early linkage maps from HHMI, Salt Lake City included many polymorphisms assayed by Southern blot hybridization on genomic DNA (for examples, see [35–39]). Along with the markers from the CRI maps, the loci included therefore lend themselves particularly well to studies such as allele loss from tumours, to which PCR cannot be straightforwardly applied. Many of the probes developed in this work have been distributed widely, and many have been deposited with the UK-HGMP and other probe banks (see below, and Appendixes IV and V for addresses).

The NIH/CEPH maps provided, in many cases, a series of maps which integrated information from previous studies into a composite map [40]. Another particular benefit of most of these maps is that many markers that cannot be placed unambiguously relative to the framework markers do nevertheless appear, with an indication of the probable location. Thus, for example, in the NIH/CEPH map of chromosome 1, D1S51 has not been placed relative to D1S74, but its location somewhere between D1S103 and D1S8 has been indicated. For some chromosomes, such as 18 and 21, this preference for comprehensive inclusion was not so thoroughly adopted, instead showing only those loci for which placement was supported at long odds; this strategy does give a well-supported framework map, but leaves many useful markers unplaced. Like the CEPH consortium maps, these maps derive from data contributed by many different groups, and thus include a wide variety of marker types.

5.2.2 More maps to come

Many more maps, at finer levels of resolution, will become available in the near future, although it is likely that the Génethon dinucleotide maps will

remain the 'gold standard' for linkage mapping for some time to come. There are already many published maps covering individual chromosomes and subregions, and more will be published—mainly in those regions where the location of a human genetic disorder provides the incentive for detailed genetic mapping. The reports of single chromosome workshops are also useful in providing up-to-date detailed mapping information.

5.2.3 From published map to useful polymorphism

In most cases the transition from a polymorphism recorded on a published map to a system that can be typed at the laboratory bench is relatively straightforward. For polymorphisms (such as dinucleotide repeats) that can be detected by PCR, primer sequences and PCR conditions can be found in the original locus description and/or in GDB. Polymorphisms typed by Southern blot hybridization will need a hybridization probe, which can be obtained either from the originator or from DNA probe banks such as the ATCC/NIH Repository (American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852-1776, USA; Tel.: (+1 301) 881-2600, Fax (+1 301) 770-2587) or the UK HGMP Resource Centre (UK Human Genome Mapping Project, DNA Probe Bank, Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ, UK; Tel.: (+44 181) 869 3446, Fax (+44 181) 869 3807). Since these have the facilities for large-scale distribution of probes, they are usually much quicker at responding to requests for probes than the originator.

5.3 Methods for identifying new polymorphisms

Where new polymorphisms need to be defined, there are several different approaches, depending on whether the starting point is a single predefined clone or sequence (see Section 5.3.1) or a large collection of DNA sequences such as a genomic library (Section 5.3.2). Methods useful in each situation will be considered in turn.

5.3.1 Starting with a defined clone or sequence

In this situation, the sequence of interest may be a candidate gene, or a cosmid, phage or other cloned sequence that is of interest primarily by virtue of its location. In either case, the priority is usually to find as much variation as possible within the given segment. For the larger cloned segments such as

cosmids or (particularly) YACs, it may be worth trying to identify tandem repeat regions (see Section 5.3.2.2). Where sequence information is available, a preliminary inspection for simple-sequence regions—including the polyadenylate tracts of retroposons—can rapidly identify potential polymorphic sites.

In the absence of tandemly repetitive DNA, there are several methods for identifying substitutional polymorphism within the defined region. Discussed below are identification of RFLPs by Southern blot hybridization (Section 5.3.1.1), and methods that require sequence information, including direct sequence analysis and single-stranded conformational polymorphism (SSCP) analysis (Sections 5.3.1.2–5.3.1.4).

Direct use of sequenced genes: intron–exon structure
One particularly useful resource when dealing with candidate genes is the sequence information from that gene lodged in a sequence database. From the sequence one can construct primers for PCR and generate amplified segments which can be screened for polymorphism, without the need to have a clone from the gene. There are a few caveats, however, arising from the fact that most gene sequences in the databases are cDNA sequences; thus primer–primer distances in the cDNA may not match those amplifiable from genomic DNA because of the presence of introns (Fig. 5.2). Where the intron–exon structure is known, and particularly where genomic DNA sequence is available, primer design and amplification of selected segments can be aimed specifically at the noncoding parts (introns, 3' untranslated and 5' untranslated) of the gene, where more polymorphism is likely than in the coding region [41,42]. In the (4-exon) example shown in Fig. 5.2, segments 1 and 2, including the second intron and the 3' untranslated region, respectively, might be suitable segments to analyse for polymorphism. Furthermore, if the gene under scrutiny

is a member of a gene family, the noncoding regions may be particularly useful for obtaining locus-specific amplification. For genetic disease studies, polymorphisms present in the protein-coding regions may be of interest, even though they may be uncommon and present at lower frequency than polymorphisms in noncoding regions.

Although PCR from genomic DNA can be used to generate a fragment from a large exon that can then be used as a hybridization probe, in screening for RFLPs a (multiexon) cDNA clone will enable more genomic DNA to be scanned than a probe (cloned or made by PCR) from genomic DNA. Thus with reference to Fig. 5.2, a genomic probe from exon I would allow the assay for polymorphism only as far as the immediately flanking restriction sites (X). In this example only one restriction fragment for X would be included. Use of the full-length cDNA, however, would include four different restriction fragments in the survey for polymorphism.

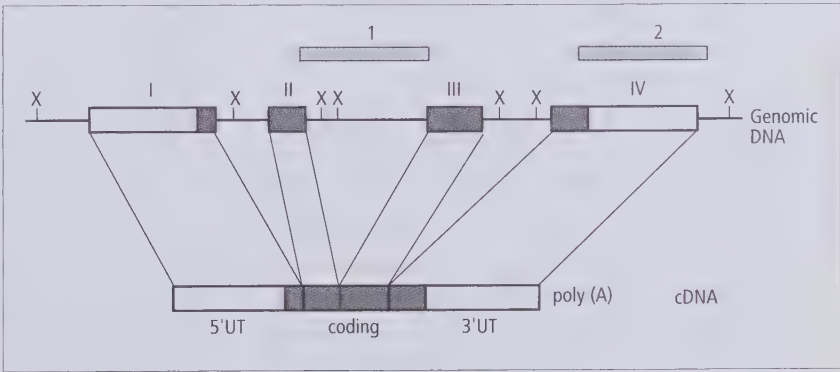
5.3.1.1 Identification of RFLPs by Southern blot hybridization

The use of locus-specific hybridization probes to identify RFLPs is long established; no knowledge of the DNA sequence is required, and the method is applicable to large or small genomic probes as well as to cDNA probes. cDNA probes have particular advantages:

- 1 the probe is very likely to recognize only a single locus (unless it is a member of a multigene family); and
- 2 if many introns are present, a disproportionately large amount of genomic DNA can be screened.

Large regions of genomic DNA, and consequently many restriction sites, can be screened using cosmid clones of genomic DNA, but the presence of dispersed repeat sequences in these probes means that locus-specific profiles will only be obtained if high copy-number elements are removed by pre-association as described in Protocol 6 [43]. Even after

Fig. 5.2 Representation of intron–exon structure in a small hypothetical gene. The greater potential for polymorphism in noncoding DNA can be exploited by preferential analysis, for example, introns (1) or 3' untranslated DNA (2).



preassociation, experience suggests variability in the quality of results obtainable between different cosmid clones.

As an alternative to removing radiolabelled high copy-number fragments from the probe by preassociation, the corresponding sites in the target DNA can be blocked on the filters by extensive prehybridization (24 h or more) in a buffer to which alkali-denatured human DNA has been added (to 0.25–0.5 mg ml⁻¹). Hybridization is carried out in the presence of similarly high concentrations of human DNA; see, for example, [44] and [45]. Prior screening for single-copy probes, by selecting clones or insert fragments which do not contain high copy-number repeats, involves more work per probe, but is more likely to give reliably interpretable results. The frequency with which RFLPs are detected appears to be the same for random and for preselected single-copy probes [46].

Screening for RFLPs is a process which could be protracted almost indefinitely, given the large number of commercially available restriction enzymes, and the possibility of finding rare variant alleles. How many people are worth screening, and which enzymes should be used?

The number of unrelated people to be screened for RFLPs depends upon the informativeness required. For even moderately informative polymorphisms (heterozygosity > 0.3), screening five unrelated individuals is unlikely to let variation go undetected; in fact, a diallelic polymorphism with heterozygosity of 0.35 has a better than 90% chance of being detected on a screen of five unrelated individuals. Less straightforward predictions can be made about the number of different enzymes that can usefully be tested. A large-scale screen of random genomic clones showed clearly that *MspI* and *TaqI* were the enzymes with the most frequently polymorphic sites, followed closely by *RsaI* [47]. This same study found that about one-third of random phage clones (insert size about 20 kb) showed polymorphism when DNA from five unrelated people was screened after digestion with six restriction enzymes [34].

5.3.1.2 Finding substitutional polymorphism: SSCP analysis

A number of sensitive and convenient methods are available for the detection of substitutional changes between DNA samples, and these have particular application in screening for mutations in genetic disease. These methods include single-stranded conformational polymorphism (SSCP) analysis [48], denaturing- or temperature-gradient gel electrophoresis (DGGE/TGGE) [49,50] (see Chapter 19)

and heteroduplex analysis [51]. Methods involving chemical [52] or RNase A cleavage [53] of mismatched duplexes are more elaborate, involving many manipulations per sample, and simpler methods are more suited to screening for polymorphism. Promising newer methods which may gain general application in screening amplified segments for polymorphism include denaturing HPLC [54] and mismatch cleavage with T4 endonuclease VII [55].

Sequence analysis is the only definitive method for demonstrating sequence variants, but other screening methods are useful in reducing the work involved in finding each variant. Once detected, the variant sequence will be defined anyway, but the usefulness of preliminary screening—here exemplified by SSCP analysis—is in identifying which individuals to sequence, and where the polymorphism is likely to be found.

Protocol 7 and the general methodology for SSCP analysis described here are modifications of established methods [48,56,57], and are direct descendants of work from Katrina Mackay and Raymond Dalglish [58] (in which these methods are applied to collagen cDNA). The main advantage of the modified approach given here is that although restriction digestion of radiolabelled fragments involves quite a lot of radioactive manipulation, a relatively large segment of DNA (as much as 1200 bp [58]) can be screened.

As an example, Fig. 5.3a shows results from the SSCP analysis of a 756-bp fragment from the human D16S309 locus [59]. DNA from six unrelated people (1–6) and a chimpanzee (C) was analysed as described above; labelled PCR products were digested with *HinfI*. Note that although in most cases the two single strands of a given size have different mobilities, the mobilities of single strands under these conditions nevertheless correlate roughly with their lengths.

Individual 1 shows a single-strand conformational variant (arrow) which appears to derive from the second largest *HinfI* fragment (184 bp). Additional variants are visible in the chimpanzee (C) lane (including a simple length variant of the third largest *HinfI* fragment—which can therefore be seen in the undenatured DNA). The position and nature of the substitutional variant for which individual 1 is heterozygous was confirmed by sequence analysis (see Fig. 5.3b).

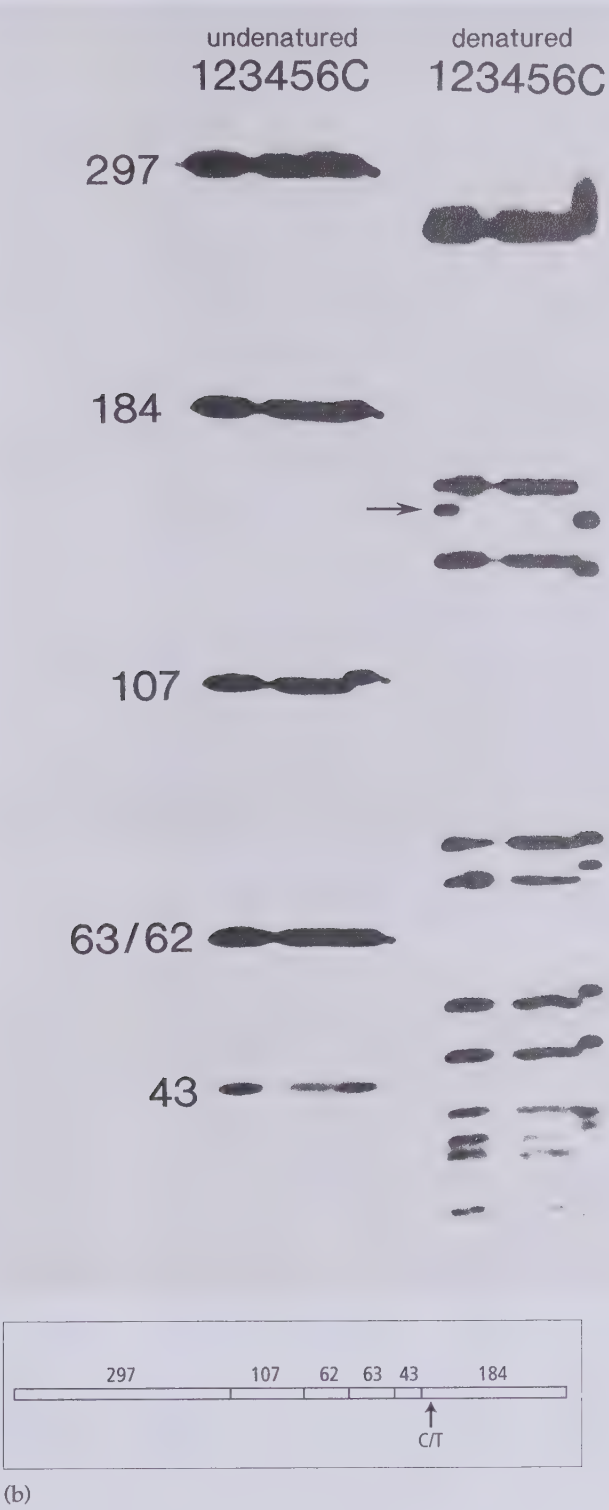
5.3.1.3 Sequence analysis

Sequence analysis of PCR-amplified segments can be used not only to verify and characterize variants observed by SSCP analysis, but also (if the segment

is short enough) as a primary method for the detection of polymorphism. Indeed, direct sequence analysis of PCR-amplified segments has been used successfully to identify a number of informative substitutional polymorphisms [60]. Several different methods for the direct sequencing of PCR products are in current use (for review, see ref. 61; see

also Chapter 22, Section 22.4). The most direct methods use the double-stranded PCR products, either in modified protocols for sequencing with T7 DNA polymerase [62,63], or in 'cycle sequencing' using *Taq* polymerase [64,65]. In each case the sequencing reactions depend on ddNTP/dNTP ratios, and thus it is important to purify the

Fig. 5.3 Polymorphism analysis.
(a) Analysis by SSCP of a 756-bp fragment from D16S309 after digestion with *HinfI*. Undenatured and denatured samples from six unrelated humans (1–6) and a chimpanzee (C) have been loaded together. A single-stranded mobility variant can be seen in individual 1 (arrow). The nature and position of this polymorphism was confirmed by sequence analysis, as shown in (b), which also shows the positions of the different *HinfI* fragments in the segment analysed.



fragment to be sequenced away from the high concentrations of dNTPs in the PCR reaction. We would generally gel purify, or precipitate with propan-2-ol; a number of proprietary methods (column or glass milk based) may also be also useful. Similarly, sequencing methods using T7 DNA polymerase may involve incorporation of radioactivity during DNA synthesis, and thus the PCR primers must also be removed before sequencing (or synthesis will be primed from both ends); this is less important with an end-labelled primer, when such unwanted products will not be radioactive.

Other methods use single-stranded templates prepared, for example, by asymmetric amplification [66] or with λ -exonuclease [67]. Other modifications, using biotinylated primers [68], or including phage promoters in extended primers [69] to sequence from RNA templates, use specially synthesized primers, and have particular applications in the repetitive sequence analysis of a single short region. First-generation automated sequence analysis using *Taq* polymerase and cycle sequencing could not be relied on to detect heterozygosity, but more recent improvements in automated sequencing reagents and software now appear to allow reliable detection of heterozygous positions.

Whichever method is used, the main priority is to produce sequence of sufficient quality, in particular to enable identification of heterozygous positions. If sequencing is being used as a primary means of detecting polymorphism, it may be useful to investigate sources of DNA in which the locus under investigation is represented only in the hemizygous state, thus avoiding the potential difficulties posed by the identification of heterozygous positions. For human X-linked loci this can simply be achieved by sequencing PCR products from males only; for autosomal loci, DNA from somatic cell hybrids could be used. Alternatively, where SSCP analysis has been used for screening, these results may identify individuals homozygous for a variant sequence.

If all else fails, clean sequence can be obtained from clones of the PCR products. It is important to remember, however, that *Taq* polymerase introduces occasional errors into the amplified DNA. When analysing clones of individual amplified molecules therefore at least two concordant sequences should be obtained before a variant sequence can be accepted as having existed in the original genomic DNA. There are many methods currently used for cloning PCR products, including those tricks used by a number of proprietary kits (see, for example, refs. 70 and 71).

If cloning without kits or other established methods, the important point to remember is that

the PCR product molecules have ends that are difficult to deal with.

1 The 5' ends (derived from the primers) will be unphosphorylated and therefore will need to be phosphorylated using T4 polynucleotide kinase and ATP before ligating into a dephosphorylated vector.

2 *Taq* polymerase frequently adds a nontemplated residue to the 3' end of PCR products, such that the products are not blunt-ended but have a single base 3' overhang. These ends can be modified by treatment with Klenow polymerase in the absence of dNTPs (which removes the overhang by its 3' to 5' exonuclease activity), followed by addition of dNTPs, allowing 'fill-in' synthesis by the polymerase; alternatively, T4 DNA polymerase can be used.

Our own practice is to avoid cloning wherever possible, but, if necessary, to 'trim' the PCR product before ligation, using unique restriction enzyme sites near the ends of the fragment. While this inevitably leads to loss of sequence from the ends, it overcomes both the problems mentioned above, by creating 5' phosphorylated ends of known structure. Digestion with proteinase K before restriction digestion has been reported to improve the efficiency of cloning PCR products by this method [72].

Having identified a sequence variant, it is usually impracticable to leave sequence analysis as the only typing method for the polymorphism; Section 5.3.1.4 describes how convenient alternative assays can be designed for such substitutional polymorphisms.

5.3.1.4 Conversion of sequence polymorphisms to convenient assays

If it is to be typed conveniently on large numbers of samples, a characterized substitutional polymorphism needs an assay other than SSCP or sequence analysis. Length polymorphisms pose no problem, in that the appropriate segment can be amplified and the alleles distinguished after gel electrophoresis (see below). Several different techniques are available for typing known substitutional sequence polymorphisms.

- Amplification across the polymorphic region, followed by hybridization with an allele-specific oligonucleotide (ASO), can be applied to many samples in parallel; PCR products can be dot-blotted in duplicate, and the filters hybridized with each ASO [73].

- An even higher turnover, and the potential for automation, can be achieved using the oligonucleotide ligation assay (OLA) [74]. This uses the discriminatory power of DNA ligase, which will only ligate two adjacent oligonucleotides if there is no mismatch between adjoining bases; the

oligonucleotides are designed such that ligation will only take place in the presence of one particular allelic form. Ligation is detected as the covalent capture of one (biotinylated) oligonucleotide by the second (immobilized) oligonucleotide in the presence of the amplified DNA. Although setting up such microtitre-plate enzyme-linked assays involves more initial work, large numbers of samples can subsequently be typed.

- Solid-phase minisequencing likewise has the potential for convenient typing of very large numbers of samples. A single primer adjacent to the polymorphic base is used to prime synthesis, with PCR product from the individual to be typed as the template. The identity of the first nucleotide incorporated gives the genotype [75,76].

- More appropriate for the typing of smaller numbers of samples are methods which can be implemented with less initial work. Methods that use oligonucleotide primers specifically designed to prime PCR selectively from one allelic form are simple to design (amplification refractory mutation system (ARMS) [77–79], PCR amplification of specific alleles (PASA)), but require two reactions per individual.

- Simpler still are systems that take advantage of the sequence specificity of restriction enzymes (see, for example, refs 42, 59 and 80). This method has the advantage that diploid genotypes can be obtained from a one-tube test, and it is the most popular method in our laboratory. Its implementation can be illustrated by an example of a substitutional polymorphism in the DNA flanking minisatellite MS31 [80] (see Case Study 5.1).

5.3.2 Starting with large DNA collections

The previous section on 'DIY' polymorphisms dealt with the pursuit of variation within a small, predefined region. This section deals with a more general situation, in which a collection of clones or DNA fragments is first sifted to find those most likely to show polymorphism. In general terms, this will most frequently involve the screening of cloned libraries by hybridization for tandem-repetitive regions.

The usefulness of the polymorphisms derived from a DNA collection will to a large extent depend upon the quality and appropriateness of the starting material. Given the large number of highly informative polymorphisms already defined, it will only be in specialized projects—for example, attempts to isolate a particular class of tandem repeats—that isolating loci at random from the human genome will be an efficient way to proceed.

Analysis of substitutional polymorphism in the DNA flanking minisatellite MS31 [80]

Figure 5.4a shows the sequence in the vicinity of the polymorphic base, which can be G or C at position –220. If the polymorphism results in the creation or destruction of a site for a commercially available restriction enzyme, the assay for the site is, in principle, very simple. In this example, the G allele can be cleaved with *HgaI* (recognition sequence 5'GCGTC3'); this site is absent if the –220 base is C. In principle therefore this polymorphism could be assayed by amplification of a segment spanning the –220 position, followed by cleavage with *HgaI*.

In fact, while a number of such systems have been designed [42,59,80], this solution was not used in this case. A simple 'amplify and cleave' assay is not appropriate if:

- there is no commercially available enzyme for which the recognition sequence is altered by the polymorphism;
- an enzyme site is altered, but the enzyme in question also cuts very frequently in the surrounding DNA, such that the allelic forms cannot be resolved among the small fragments; or
- the enzyme is prohibitively expensive.

HgaI is indeed very expensive and so a strategy was adopted that used a deliberately mismatched primer (31Rsal). Amplification between this primer and a flanking primer (to the left (not shown) in Fig. 5.4b) alters the sequence next to the polymorphic base such that the PCR products now contain a site for *RsaI* (5'GTAC3') in products amplified from a –220G allele (Fig. 5.4b). Thus by using PCR with a deliberately mismatched primer to force a change (bold) into the neighbouring DNA sequence, a restriction site polymorphism can be engineered into the PCR products, resulting in a simple two-step, one-tube assay for genotyping at this locus (Protocol 8) [80]. A simple 'amplify and cut' assay for substitutional polymorphisms can therefore be used either when the substitutional variation changes a restriction site, or where a mismatched primer can be used to engineer such a site. Protocol 8 outlines the general methods used to type this class of polymorphism.

Case Study 5.1

In less well-characterized organisms, such as non-human, non-murine vertebrates, it may well be worth starting with a library representative of the whole genome; in human genetics, by contrast, interest will usually be directed to a single chromosome or subchromosomal region. The details of library preparation in these cases are beyond the scope of this discussion (see a general cloning manual such as ref. 81) but the following general routes to 'subgenomic' libraries are worth noting.

5.3.2.1 Starting points for subgenomic libraries

Somatic cell hybrids A somatic cell hybrid containing the chromosome of interest as its only human

component can be a useful starting point for a chromosome-specific library (Chapter 14). The human chromosome can be separated from the rodent background by flow-sorting (see below, and Chapter 12) or a library can be constructed directly from the hybrid, and clones containing human DNA selected by hybridization with total human genomic DNA (see, for example, [82]). Labelling total genomic DNA for use as a hybridization probe will result in signals only from high copy-number elements; Alu elements are absent from rodents, but abundant in human DNA (about 1 element every 5–6 kb [83]) and thus several should be present in nearly all human cosmid clones.

Flow-sorted chromosomes Flow-sorting can be used to enrich for a chromosome of interest either directly from human cells, or from single-chromosome somatic cell hybrids (see Chapters 12 and 14). Flow-sorted DNA can then be used to construct a library. Much less flow-sorted DNA is needed to produce DNA collections amplified by PCR, and procedures such as DOP-PCR, which use a universal driver oligonucleotide [84,85] (see also Protocol 52, Chapter 10), also give libraries that are easily renewable by PCR.

Alu-PCR The absence of Alu elements from rodent DNA forms the basis of a convenient method for specifically amplifying human DNA fragments from rodent–human somatic cell hybrids. This method uses consensus primers to amplify between adjacent Alu elements; the resulting fragments will be highly enriched for human sequences. While this and other methods (such as DOP-PCR) for producing chromosome ‘paints’ (see Chapters 10 and 11) will not usually produce a representative collection of fragments, they can form a useful starting point for the isolation of chromosome-specific polymorphisms.

Radiation hybrids Defined regions smaller than a whole chromosome can be isolated after fragmentation of chromosomes by X-irradiation (Chapter 14). Human DNA can be recovered from these hybrids by the hybridization and PCR methods outlined above for whole chromosomes.

Microdissection libraries A more direct, but more demanding approach is to isolate the region of interest by dissecting it out from chromosome spreads (Chapter 11). In this case the amount of DNA recovered is small, and PCR methods are required to prepare a library.

YACs, YAC contigs, cosmid contigs Relatively small fractions of the genome, such as those forming clone contigs between flanking markers, may still contain many hundreds of kilobases of DNA, and a library-based approach to the isolation of polymorphisms from such regions is appropriate.

Although the general methods outlined in Section 5.3.1 could in principle be applied to each clone in a library, methods with higher turnover are required in library screening. Hybridization screening for tandem repeats, as described in the following sections, remains the simplest method for identifying sequences likely to display polymorphism.

5.3.2.2 Tandem repeat variability: VNTRs

Tandem repeats form a considerable fraction of many genomes, and at many loci the number of repeats shows allelic variation. This variability in tandem repeat number occurs on several different scales, from mononucleotide repeats to satellite arrays. For this reason the acronym VNTR (variable number of tandem repeats) strictly applies to all these classes of locus, not just to the medium-sized arrays (‘minisatellites’) to which the term was first applied [44]. Other terms, such as ‘microsatellite’, ‘minisatellite’, ‘dinucleotide repeat’, etc., give more detail about the type of locus involved, but say nothing about variability; indeed, many minisatellite and microsatellite loci show no detectable polymorphism [44,86,87].

Dinucleotide repeats, and in particular (AC)_n repeats, combine the general advantages of abundance, even distribution and frequent, highly informative polymorphism [88–90]. For most projects,

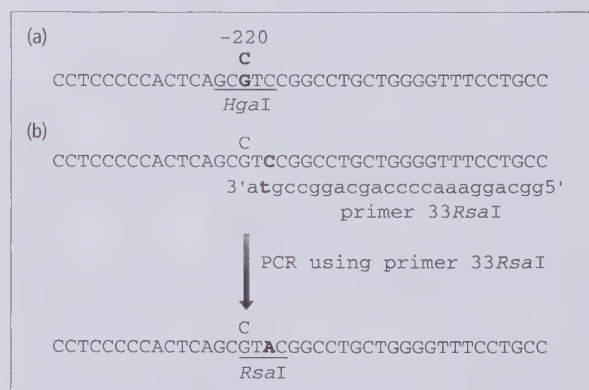


Fig. 5.4 Analysis of a substitutional polymorphism. (a) Substitutional polymorphism (C/G) at position -220 of MS31. There is a site for *HgaI* (CGGTC) in the -220G allele which is absent from the -220C form. (b) Assay for the -220 polymorphism by PCR with a mismatched primer (31RsaI). By forcing a base change (bold), this primer creates a new content for the polymorphism, such that the substitution now creates or destroys a site for *RsaI* (GTAC).

these loci would represent the most efficient target for rapid isolation of new polymorphisms. Other tandem repeat types may also have particular uses: thus where $(AC)_n$ dinucleotide repeat loci have already been isolated to saturation, other abundant dinucleotide (e.g. $(AG)_n$) repeat loci may be useful. Where the target locus is a genetic disorder that may be due to tandem repeat instability, isolation of trinucleotide repeat loci not only provides genetic markers but may provide a more direct route to the gene [91].

Sequence analysis and primer design At short repeat loci typed by PCR, analysis of the sequence flanking the repeats is necessary before primer design. The single simplest check is to search the sequence databases using only the (nonrepetitive) flanking DNA. In humans, the most frequent questions sequence analysis needs to address include:

- is the sequence new? (or has someone else already submitted it the database?);
- where do the repeats begin and end? (there may be variant repeats at one end of the array—which may themselves vary in number: don't include these in the primer);
- is there a dispersed repeat element in the sequence? (see below).

A number of different programs are available for the computer-assisted design of oligonucleotide primers, of which the most widely used are probably PRIMER (from the Whitehead Institute) or OLIGO [92]. PRIMER can be obtained via anonymous ftp (for details, see [93]). Our laboratory has not used these programs, but instead followed simple rules for the design of primers 'by eye'.

- 1 Make the primers as GC-rich as possible (but see point 3).
- 2 Avoid runs of simple repeats, such as AGAGAGAG, particularly near the 3' end of each primer.
- 3 Match the two primers either side of the array for melting temperature (see '4 + 2' rule below); use the length and composition to aim for temperatures within a range (about 40–75°C) at which *Taq*

polymerase will extend efficiently and specifically. For sequences of representative composition, 18–22mer primers work well.

- 4 Avoid dispersed repeat consensus sequences (see below).
- 5 Within these constraints, put the primers as close to the repeats as possible.

In our experience, a simple '4+2' rule has proved useful in primer design; count 4°C for every G/C and 2°C for every A/T. While the sum so obtained is usually rather less than can be used as an annealing temperature in PCR, matching the sums for two primers has given primer pairs which work well together.

Typing of the simple repeat locus d3s1749 (=wg1e7, EMBL accession no. x74780 [111])

In this sequence (Fig. 5.5), a block of (AAGG) repeats is found immediately adjacent to an Alu element; the primer sequences chosen are underlined.

One solution is simply to end-label the nonAlu primer (2792), so that even if the Alu primer (2744) primes from multiple sites in the genome, only locus-specific products will show up on the autoradiograph. This has the economic advantage that a consensus Alu primer can be designed, which could subsequently be applied to other similar loci. Another solution, which will minimize the yield of other products, is to design the Alu primer such that it is as divergent as possible from the consensus Alu element sequence. We have found the sequence of Bains [136] useful for this purpose. In Fig. 5.5, the bases which match the Bains consensus Alu element are shown in bold capitals: primer 2744 has been designed to contain a large number of mismatches, particularly towards its 3' end. In fact, this locus can be amplified from human genomic DNA (with an annealing temperature of 60°C) to give locus-specific products visible on an agarose gel, with little coamplifying DNA from other loci.

Combining these two approaches (carefully designed Alu primer with end-labelled nonAlu primer) has been very successful in our laboratory in giving clean typing at simple repeat loci associated with dispersed repeats.

Case Study 5.2

Fig. 5.5 Primer design at the Alu-associated tetranucleotide locus D3S1749. The primer 2744 has been designed to contain many differences from the consensus Alu sequence; bases matching this consensus are shown in bold capitals.

[illegible]

Coping with dispersed repeats In order to type short repeat loci arising from the polyadenylate tracts of dispersed repeats (such as Alu elements) (without giving a product so large that closely spaced alleles cannot be resolved), one of the PCR primers must be placed within the dispersed repeat sequence. Locus-specific profiles can reliably be obtained, however, if care is taken over primer design. The general principles are illustrated in Case Study 5.2 with reference to the human simple repeat locus D3S1749 shown in Fig. 5.5.

5.3.2.3 Screening for polymorphism: general considerations

It is an unfortunate fact that having identified and subcloned a tandemly repeated region, it turns out that polymorphism is not absolutely guaranteed. At both dinucleotide and minisatellite loci, the same simple rule appears to hold—the longer, the better [87,93]. Nevertheless, about a half of the 4–9 kb minisatellite arrays examined seem to be monomorphic [86,95]. However, after careful selection of dinucleotide arrays (containing more than 12 repeats) nearly all loci tested were polymorphic [12]. Dinucleotide repeats containing perfect repeats appear to be more reliably polymorphic than loci containing variant repeats [89]. By contrast, interspersed variant repeats are a common feature of even the most highly polymorphic minisatellite loci [96]. Among triplet and tetramer repeat loci, the tetramer repeats appear to be generally longer and more variable in human DNA.

Screening four or five unrelated individuals should be enough to identify all polymorphisms—except the relatively uninformative ones (see Section 5.3.1.1). However, where the study is geared towards the analysis of a particular pedigree (either an extended cross between two inbred strains, a single large human pedigree, or recombinant inbred mouse strains), it is not particularly important whether the locus is polymorphic in the general population, but crucial that the founders of the pedigrees are informative. It therefore makes most sense in this situation to test for polymorphism using those individuals from which segregation is being scored.

Is the polymorphism new? It can be particularly irritating to spend time isolating and characterizing a new polymorphic locus, only to discover that it has been isolated independently elsewhere. In the interests of efficiency and equanimity it is therefore helpful to detect such duplications at an early stage. For PCR-amplified polymorphisms, the DNA sequence serves to identify the locus unambigu-

ously; thus a search of the Genbank/EMBL sequence databases should be enough to prevent unnecessary duplication of effort. In humans, GDB is also a useful source of information on existing polymorphisms. These considerations make it particularly important for each worker to make sure that data concerning any new polymorphism are submitted to the relevant databases (if only to assert priority in its discovery!).

Minisatellites, which can be analysed without the need for sequence data, pose more of a problem. Within a laboratory, test panels of DNA from standard individuals can be used to test for repeat isolates from a known locus. Standardization between laboratories requires DNA samples to which all concerned have access [97]. A case of duplication has nevertheless come to light surprisingly late [98].

5.3.2.4 Mononucleotide repeats

The number of A residues composing the polyadenylate tract of retroposons—and, in particular, the human Alu elements—can show polymorphism [99]. This observation in itself suggests a copious supply of polymorphism which can be drawn on; although most polymorphism seems to have been defined in Alu tails containing diverged tetramer and other repeats (e.g. AAAT, AAAG, etc., [93,100]), the number of A residues in simple (A)_n blocks may also be usefully polymorphic. Alu elements occur (on average) about every 5–6 kb in human DNA, and L1 elements, which are also polyadenylated, occur about every 150 kb [101]. Thus, finding potentially polymorphic tracts is straightforward enough—screening with total human genomic DNA as a hybridization probe will identify high copy-number elements, among which the Alu elements usually give the strongest hybridization signals. One of the primers used for PCR will necessarily come from a high copy-number repeat, and thus primer design will help to obtain locus-specific products (see Section 5.3.2.2 above). For typing conditions see Section 5.3.2.5.

5.3.2.5 Dinucleotide repeat loci

Sequences containing dinucleotide repeats have been identified by hybridization screening with end-labelled repeat oligonucleotides, or with a (long) synthetic poly(AC) probe [12,82,89]. If small insert libraries (in plasmid or M13 vectors) are screened, the sequence around the repeat array can be determined without subcloning. Cosmids have the disadvantage that the repeat-containing region needs to be subcloned (or adjacent regions isolated by PCR methods [102]) before the sequence is

determined, but have the advantage that the original cosmid clone can subsequently be used for physical localization (by FISH, see Chapter 9). Methods involving preselection by hybridization may be useful in enriching for repeat-containing fragments [103,104]. If large numbers of repeat loci of the same repeat sequence are to be isolated, strategies using specially synthesized sequencing primers may speed up the process [105].

Typing dinucleotide repeats In typing dinucleotide and other simple repeat arrays by PCR two opposing considerations apply as more and more cycles of PCR are done. The amount of product rises as amplification proceeds, but so also does the tendency to form artefacts containing different numbers of simple repeats from the template DNA. The tendency to produce these artefacts is generally rather less pronounced for tri- and tetranucleotide arrays [106]. Even within the class of dinucleotides, the load of artefactual products can differ greatly between different loci. It is therefore recommended that conditions (particularly of cycle number) should be explored with every new system. Protocol 9 is intended to serve as a guide to this process of optimization.

5.3.2.6 Tri- and tetranucleotide repeats

The polymorphism shown by arrays containing trinucleotide and tetranucleotide repeat motifs is of considerable interest, not only because they appear to be an abundant, relatively untapped and reliably typed source of polymorphism, but also because of the direct involvement of polymorphic trinucleotide repeats in the pathogenesis of some human genetic disorders. Such disorders include fragile X-linked mental retardation, myotonic dystrophy, spinal and bulbar muscular atrophy (Kennedy's disease) [107], and Huntington's disease [108].

Such disease loci, however, are simply those short repeat loci at which array expansion causes a disease phenotype; there are also many other loci at which repeat copy number has no apparent phenotypic consequence, and which form an abundant source of polymorphism [106]. The advantages of these 'simple tandem repeat' loci are that they can be typed by PCR with little or no tendency to artefactual 'slippage' products, and that they are widely dispersed throughout the genome. Additionally (at least in the case of triplet repeats), they may even present themselves as candidates for the pathogenesis of genetic disorders. For this reason some studies have identified triplet repeat loci specifically from cDNA libraries [109,110].

Trinucleotide and tetranucleotide repeat loci are

strongly associated with retroposon tails; this is particularly true of purine-rich (e.g. AAAG) repeats, which commonly arise from Alu and other retroposons. Thus careful attention to primer design will be rewarded by better locus specificity (see Sequence analysis and primer design, above); different methods for typing different loci are therefore also appropriate.

Isolating trinucleotide and tetranucleotide repeat loci Libraries can be most simply screened by hybridization using end-labelled synthetic repeat oligonucleotides. Most successful studies have used 24–32mer oligonucleotides, such as (AAAT)₆ or (CTG)₁₀ [93,103,111,112]. As with screening for dinucleotide repeats, a small-insert library minimizes the work involved in characterizing each positively hybridizing clone, but the use of cosmids has the particular advantage that, in parallel with genetic mapping, corresponding cytogenetic placement can be made using FISH (Chapter 9). Pre-enrichment methods have also been successful [104,112]. The DNA sequence immediately flanking the repeats in positively hybridizing clones can be determined either after subcloning, or possibly by more efficient PCR-based methods [102].

Typing trinucleotide and tetranucleotide repeat loci The method appropriate for a given locus depends largely on the size of the PCR product and the number and spacing of the alleles. Thus a locus with, for example, six alleles each separated by 4 bp over the 250–270 bp range, should be typed by end-labelling a primer and running PCR products on a polyacrylamide gel (see Section 5.3.2.5 above). Although an ethidium-stained 3% agarose gel would show the polymorphism, and give an indication of the segregation in a pedigree, it would not resolve closely spaced alleles reliably enough to give dependable results. By contrast, if preliminary work had shown that a locus had only two common alleles of about 260 bp and 280 bp, these widely spaced alleles could be easily resolved by typing on a 3% agarose gel. This, of course, assumes that the locus can be amplified by PCR to give locus-specific products. At some loci, however, PCR is primed at one end from within a retroposon, and end-labelling can be necessary to show locus-specific products (see Section 5.3.2.2).

Loci with shorter alleles—for example, in the range 70–120 bp—can be typed on denaturing polyacrylamide gels after PCR with an end-labelled primer. Alternatively, experience from our laboratory suggests that (at those loci at which locus-specific products can be detected on ethidium-stained

agarose gels) sufficient resolution can be afforded by a high percentage agarose gel. Thus, for example, a triplet repeat locus with alleles of 72 bp and 75 bp can be reliably typed on a 4% (NuSieve) agarose gel.

Heteroduplex formation Unlike (single-stranded) PCR products resolved on a denaturing polyacrylamide gel, double-stranded products will at many loci give not two but three bands from heterozygotes after nondenaturing gel electrophoresis. This third band, which usually migrates more slowly than either of the homoduplex fragments, is due to the annealing during the final PCR cycle of one strand from each allele to form a heteroduplex. Although this phenomenon does not usually cause ambiguities in typing, it may cause some concern if its origin is not understood. In practice, preliminary work with denaturing polyacrylamide gels should clarify the origin of the 'extra' band in heterozygotes.

5.3.2.7 Minisatellites

Minisatellites—repeat arrays typically in the range 1–30 kb, and composed of 8- to 100-bp repeats—can show extraordinary levels of variability [112]. They may be typed as single loci, using a locus-specific probe at high stringency, or, using a tandemly repeated probe at low stringency, many hypervariable loci can be typed simultaneously. Multi-locus methodology will be discussed in the chapter on DNA typing (Chapter 6); here I will deal with the development of single-locus minisatellite probes.

The general advantages of minisatellite loci are the very high levels of informativeness, and the fact that length polymorphism can be assayed as an RFLP system, but nearly independently of the restriction enzyme. Thus a number of different loci could be typed by repetitively probing blots of genomic DNA cleaved with (for example) *Mbo*I, or any other enzyme which does not cleave within the repeated alleles. This facility for typing the same samples at many different loci without starting

afresh for each locus may have benefits in screening, for example, for allele losses from tumours. The main disadvantage is that relatively large amounts of genomic DNA are required. PCR typing can reliably be applied to some minisatellite loci [113,114], but the difficulty of amplifying the longest alleles limits the usefulness of PCR typing [115]. The high mutation rates observed at the most informative loci [116,117] may also cause some difficulty in family studies.

Many human minisatellite loci have already been isolated from genomic libraries, by a variety of techniques [44,45,86,95,117–120]. As ever, the best way to get hold of a useful locus is to find out about one that has already been isolated. Because of their unusual distribution, although minisatellites may be of doubtful value in some studies, where the gene of interest maps to a subtelomeric location they may represent the most valuable type of marker [4,5].

The simplest way to isolate minisatellites from a library is by hybridization screening. Listed in Table 5.2 are probes that have been found useful in identifying minisatellites by screening human genomic DNA libraries under low stringency conditions; similar approaches have been adopted in isolating locus-specific minisatellite probes from nonhuman species [121,122]. In general, probes that detect multiple hypervariable loci on low stringency hybridization of genomic DNA ('DNA fingerprinting') can be used under the same conditions to screen libraries for clones containing those hypervariable loci; some of these probes are discussed further in Chapter 6.

Typing minisatellite loci Unlike microsatellites, minisatellites can be typed and evaluated without the need for sequence information. A cross-hybridizing (and thus presumably tandemly repeated) fragment from the clone is labelled and used to probe genomic DNA at high stringency. Most minisatellites we have isolated give good, locus-specific signals after washing filters in 0.1 × SSC, 0.01% SDS at 65 °C. With

Probe	Type	Reference(s)
33.6	(a)	[86,120]
33.15	(a)	[86,120]
α-globin 3' HVR	(a)	[86,123]
MS1	(a)	[86]
Various	(b)	[44,45]
(CAC) _n	(b)	[121]
'STR' probes	(c)	[98,118]

Table 5.2 Probes used for screening for minisatellites.

Probe type: (a) human genomic DNA clone; (b) synthetic oligonucleotide; (c) synthetic tandem repeat array.

some probes, whereas the main locus gives the predominant signal, there may still be a residual signal from other, cross-hybridizing loci [119]. Another feature peculiar to Southern blot hybridization at these loci is that the hybridization signal obtained on probing with a tandemly repetitive probe will be proportional to the number of repeats in an allele, and thus longer alleles will give more signal than shorter alleles.

5.4 Multilocus methodology

5.4.1 DNA fingerprinting: general principles

This section gives an overview of systems that do not seek to analyse a single locus in isolation, but instead scan a number of unlinked loci with a single test.

The basic principle underlying DNA fingerprinting is that many tandemly repetitive probes will cross-hybridize with other, noncognate loci under conditions of low stringency. The most useful probes are those which cross-hybridize with a large number of other loci, many of which are themselves highly polymorphic. The combined resolving power of analysing many hypervariable loci in a single profile (the 'DNA fingerprint') makes the result individual-specific [94,123,124]—with the obvious exception of identical twins [125]. DNA fingerprinting is discussed in detail in Chapter 6.

5.4.2 Random amplified polymorphic DNA PCR

Random amplified polymorphic DNA (RAPD)-PCR scans the genome for a number of arbitrarily primed polymorphic loci [126,127]; the method requires no prior knowledge of polymorphism in the genome in question, and, as in DNA fingerprinting, the loci detected are (at least initially) anonymous, in the sense that further work is required before a locus can be analysed singly. RAPD-PCR is chiefly of use for analysing genomes where few polymorphic loci have been characterized; in humans and other well-mapped organisms, locus-specific polymorphisms of known location are preferable. Although in many cases RAPD-PCR gives a simple pattern of bands that segregate in a (dominant) mendelian fashion, a particular primer may not give a genetically correct result in a particular species. One study using RAPD-PCR in baboon families, for example, gave high frequencies of non-parental bands in offspring [128].

5.4.3 Genomic mismatch scanning

This technique identifies regions of identity between

DNA samples. The method involves the enzymatic methylation of one DNA sample, followed by solution hybridization and selective destruction (with restriction enzymes) of duplexes derived exclusively from one sample or the other. Mismatched duplexes are then destroyed using the MthLS mismatch repair proteins from *Escherichia coli*: Surviving hemi-methylated hybrid DNA should come from those fragments in which the two samples do not differ in DNA sequence, thus identifying regions of the genome held in common between the samples. This technique has been demonstrated in a (yeast) model system [129], and is a promising approach to the isolation of regions identical by descent from, for example, humans sharing a genetic disease allele.

The difficulties in extending this methodology to the human genome are chiefly due to the larger size of the genome, and the presence of high copy-number repeated DNA. Despite these potential problems, the technique is very powerful in principle, has been demonstrated to be of practical utility in yeast, and may yet have important applications in the analysis of larger genomes.

5.4.4 Representational difference analysis

The recently described technique of representational difference analysis (RDA) uses hybridization and differential PCR to identify restriction fragments present in one sample but missing from another [130]. This could arise either by deletion of the segment in question, for example, in a sample of tumour DNA, or by polymorphism in the length of a restriction fragment, such that it no longer appears in the size fraction selected. RDA has a wide potential for the specific isolation of polymorphic or deleted fragments, where the genomic location of a genetic lesion is not known by other means, as well as potential applications in the analysis of differential gene expression and pathogen identification. It has been used with spectacular success in the identification of a region of homozygous deletion in a pancreatic cancer [131].

5.4.5 AFLP scanning

The display of large numbers of restriction fragments in the AFLP technique [132] is a method of limited usefulness in species (such as humans) in which substitutional heterozygosity is infrequent. In species with high frequencies of RFLPs, and in particular in plants, it can be extremely powerful in 'scanning' a large number of loci for linkage to a trait of interest.

Protocol 6 Pre-association of dispersed repeat elements

Materials

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

- cosmid DNA for labelling
- oligonucleotide primers
- labelling buffers
- [α - 32 P]dCTP (3000 Ci mmol⁻¹, 10 mCi ml⁻¹)
- stop solution: 20 mM NaCl, 20 mM Tris-HCl (pH 7.5), 0.25% SDS
- high molecular weight herring sperm DNA (3 mg ml⁻¹)
- 2 M sodium acetate, pH 5.6
- 100% ethanol
- 80% ethanol
- 0.1 M NaCl
- alkali-denatured human DNA (approx. 8 mg ml⁻¹) (see auxiliary protocol)

Method

- 1 Label the probe by random oligonucleotide priming (see [133]). Label 10–20 ng of cosmid DNA, using 5 μ l [α - 32 P]dCTP (3000 Ci mmol⁻¹, 10 mCi ml⁻¹).
- 2 Recover the labelled fragments from the unincorporated dCTP: to a 30- μ l reaction add: 70 μ l stop solution, 30 μ l herring sperm DNA, 30 μ l 2 M sodium acetate, 425 μ l 100% ethanol.

A DNA pellet should be easily visible, in which the labelled fragments will coprecipitate. Remove the supernatant: if the labelling has worked well, approximately 50% or more of the radioactivity should be incorporated. Wash the pellet with 425 μ l 80% ethanol.

- 3 Dissolve the probe in 450 μ l 0.1 M NaCl; add 50 μ l (8 mg ml⁻¹) alkali-denatured human DNA (see below). DNA sheared by sonication can also be used. Incubate this mixture (at 100 °C for 1 min), and incubate at 65 °C for 10–12 h (Cot = 90–110^a).
- 4 Add the pre-associated probe to the hybridization mix: resist the temptation to boil again at this stage!

Preparation of alkali-denatured human DNA

Materials

- human DNA
- 3 M NaOH/200 mM EDTA

Standard procedures for radioisotope labelling by random oligonucleotide priming can be found in ref. 133.

^aCot values. Co, the concentration of DNA, is measured in mol (of nucleotide) per litre. Thus, in the example above, the DNA concentration is 400 μ g in 500 μ l, or 0.8 mg ml⁻¹, equivalent to 0.8 g l⁻¹; the Co is therefore 0.8/318, or about 0.0025 mol l⁻¹ (the average deoxyribonucleotide residue in DNA has a relative molecular mass of about 318). Cot values are given in mol seconds per litre (mol s l⁻¹). Incubation for 10 h (= 36 000 s) in the example above thus gives a Cot of about 90.

Method

- 1 To 9 vols human (usually placental) DNA, add 1 vol. 3 M NaOH/200 mM EDTA; incubate at 100 °C for 5 min. This violent treatment serves to denature the DNA, but also to shear it into smaller fragments. Neutralize the alkali with HCl, precipitate with ethanol, wash with 80% ethanol, vacuum dry and redissolve in water to 8–10 mg ml⁻¹.

Troubleshooting

No signal visible

A human DNA probe used in a hybridization with human DNA should give a signal, although there is variation in signal intensity between different probes.

- *Check the Southern blot filters using a single-copy probe known to give good signals.*
- *Check the incorporation at the labelling stage.*
- *Increase the amount of genomic DNA per lane, say from 5 µg to 10 µg.*

Smeared signal

If repeat sequences are incompletely suppressed, hybridization of large numbers of genomic fragments with the probe will give a smear of signal in each lane.

- *Add denatured human DNA to the hybridization (see discussion below).*
- *If necessary (although more work), use smaller restriction fragments isolated from the cosmid.*

Protocol 7 SSCP analysis for polymorphism

Materials

- genomic DNA
- PCR buffer (dNTPs < 0.1 mM; final concentration, see discussion)
- PCR primers
- Taq DNA polymerase
- paraffin oil
- 7.5 M ammonium acetate
- propan-2-ol
- 80% ethanol
- restriction enzyme buffer(s)

- restriction enzyme(s)
- stop solution: 20 mM EDTA/0.2% SDS
- loading buffer: 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol
- equipment and materials for polyacrylamide gel electrophoresis (6% polyacrylamide gel containing 10% glycerol in 1×TBE buffer)

Method

- 1 PCR-amplify the fragment to be analysed from four or five unrelated individuals (see ref. 73 for technique). Check on a gel that the amplification is clean—that is, it has given a single product of the expected size. If there are significant coamplifying fragments, the correct fragment can be gel-purified at this stage.
- 2 Use 0.1 ng of the recovered DNA to seed a PCR reaction containing 0.5 µl [α - 32 P]dCTP (3000 Ci mmol⁻¹, 10 mCi ml⁻¹), in a total volume of 50 µl. Do 16 cycles under standard cycle conditions.
- 3 Remove the (radioactive) aqueous layer from under the paraffin oil; precipitate the amplified DNA by adding 25 µl of 7.5 M ammonium acetate and 150 µl propan-2-ol. Centrifuge for 10 min; wash the pellet twice in 150 µl 80% ethanol. Redissolve in 20 µl water.
- 4 Digest samples containing 10–50 c.p.s. (usually about 2 µl) with the appropriate restriction enzyme, in a total volume of 10 µl. Stop the reaction by adding 10 µl of 20 mM EDTA (pH 8.0)/0.2% SDS.
- 5 Mix 10 µl of each restriction digest with 10 µl loading buffer; split this mixture between two tubes.
- 6 Run the samples on a 6% polyacrylamide gel containing 10% glycerol in 1×TBE buffer. Load 5 µl per lane from each set of digests (a) undenatured, from the first of the duplicate tubes in step 5, and (b) after denaturing at 85–100 °C for 2 min, from the second tube. Run the gel at a constant, low power (such as 30 W) to keep the temperature from rising too high; running the gel in a cold-room allows the use of a reasonably high voltage without overheating. The bromophenol blue migrates (approximately) with double-stranded fragments of about 50 bp under these conditions.
- 7 Dry the gel (no need to fix) and expose to X-ray film. Characterize any variants observed by sequence analysis (see below).

Discussion The re-amplification of the fragment in the presence of [α - 32 P]dCTP (step 2) uses a PCR buffer containing relatively low dNTP concentrations, so that the incorporation of 32 P into the PCR product is correspondingly efficient. We generally use a modified PCR mix which differs from our usual buffer in containing a (final) concentration of only 0.1 mM of each dNTP. While the method for recovery of the 32 P-labelled PCR product (step 3) is not critical, experience suggests that precipitation with propan-2-ol, followed by 80% ethanol washes,

results in less contamination of the product with unincorporated [α - ^{32}P]dCTP than, for example, ethanol precipitation.

The choice of restriction enzyme(s) will be dictated by the sequence. In general, best resolution of single-stranded conformational variants is given by fragments in the 50- to 300-bp range. In most cases it is possible to find two different digests that between them will ensure that any part of the amplified sequence is represented on at least one fragment in this range. Loading each sample undenatured (in addition to the denatured sample for SSCP analysis) allows a check that the restriction digestions have worked, will show any RFLPs for that enzyme, and allows a side-by-side comparison which will help deduce the location of any variation seen.

In a typical experiment, ^{32}P -labelled DNA amplified from four unrelated people might be digested with two different enzymes (or combinations of enzymes), and these eight samples split and run with and without denaturation. Such an experiment would thus use 16 lanes on the gel. It is most helpful to load all similarly processed samples from different individuals in a group, and to put the native and denatured loadings of the same groups of samples side by side (see Fig. 5.3a). Gel electrophoresis can take a long time if the gel is not to be allowed to warm up; typical runs on a 50-cm gel have taken 6–7 h.

.....

Troubleshooting

No bands seen

Poor incorporation of radioactivity during step 2 may result in bands being faint or invisible. Remember that the smaller fragments will contain less radioactivity (see Fig. 5.3a).

- *Monitor incorporation by comparing the pellet with the supernatant at step 3. Use a blank control (no DNA input at step 2).*
- *Do a non-radioactive PCR at step 2; check on a gel for the appearance of the correct product.*

Too many bands

Incomplete digestion in step 4 will lead to the production of additional bands, corresponding to incomplete digestion products; this is the point of loading undenatured DNA (see discussion). If there is a clean difference in one or two bands only between some samples, remember this may be a real RFLP! Incomplete digestion will give many extra bands, which will be larger than expected.

- *Add more restriction enzyme.*
- *Digest for longer.*

Protocol 8 Typing amplified RFLPs (PCR-RFLPs)

Materials

- genomic DNA
- PCR buffer (see, e.g., ref. 115, and Appendix I)
- PCR primers
- *Taq* DNA polymerase
- paraffin oil
- restriction enzyme buffer (as per manufacturer’s instructions)
- restriction enzyme

Method

- 1** Amplify the segment to be analysed, using PCR in a volume of 10 µl per sample. Note that even in ‘oil-free’ PCR machines such a small volume will need covering with paraffin oil. If possible, include individuals of known genotype as standards.
- 2** Make a 1.5-µl restriction digestion mix, containing (per sample):
 - 3 µl restriction enzyme buffer (10×);
 - 3 units restriction enzyme;
 - distilled water to 20 µl.
- 3** Add (under oil) 20 µl of the 1.5 × mix to each 10 µl PCR. Mix and incubate at the correct temperature for the enzyme, 2–4 h.
- 4** Analyse samples on an agarose gel (2–4%, depending on the expected fragment sizes).

.....
Troubleshooting

No fragments seen

- The initial PCR may not have made enough product to analyse.
- *Check that the fragment of interest amplifies cleanly and with good yield under the conditions used.*
 - *If necessary, increase the number of PCR cycles, or change conditions, in particular annealing temperature or Mg²⁺ concentration.*

Too many fragments

- This may be due to additional products generated by the initial PCR or to incomplete digestion (see below).
- *Check undigested material on a gel. If this is more complex than expected, examine the PCR conditions to reduce nonspecific*

products. In particular, look at annealing temperature, Mg^{2+} concentration and cycle number.

- If undigested material looks correct, see ‘incomplete digestion’ below.

Incomplete digestion

Most enzymes tested in our lab tolerate the admixture of PCR buffer with no effect on specificity or efficiency. If the enzyme is clearly not digesting completely:

- Increase the amount of enzyme to 10 units per tube, and incubate for longer (overnight).
- Reduce the relative admixture of the original PCR: for example, do 5 μ l initial PCRs and add 40 μ l of a 1.12 \times restriction enzyme mix.
- Test an alternative isoschizomer (if available). Thermophilic enzymes (such as TaqI, BstUI, BsaJI) appear particularly robust.

Protocol 9 Typing dinucleotide repeat loci

Materials

- genomic DNA
- PCR buffer
- PCR primers
- T4 polynucleotide kinase
- T4 kinase buffer
- Taq DNA polymerase
- paraffin oil
- loading buffer: 95% formamide, 20 mM EDTA (pH 8.0), 0.05% bromophenol blue, 0.05% xylene cyanol
- fixative for gel: 10% methanol, 10% acetic acid
- materials and equipment for denaturing polyacrylamide gel electrophoresis

Method

- 1 Label one primer using [γ - 32 P]ATP and polynucleotide kinase; if there is a dispersed repeat at one end of the array, label the primer from the single-copy DNA. We would generally label 1.5 pmol primer per subsequent PCR reaction; for example, if 20 samples are to be typed, label 30 pmol primer in 20 μ l, and use 1 μ l labelled primer in each PCR reaction.
- 2 Use 100 ng input DNA and 1.5 pmol labelled primer (above), together with 10 pmol of the other (unlabelled) primer and 0.5 U Taq polymerase in a 10- μ l PCR. Try 18, 20, 22 cycles in initial tests.

- 3 Add 5 µl loading buffer; denature the sample by replacing in the heating block and heating to 99 °C for 2 min. Keep on ice before loading.
- 4 Run on a denaturing polyacrylamide gel; fix (10% methanol, 10% acetic acid, 10 min), dry and autoradiograph.
- 5 Titrate the number of cycles to give the best results: if the signal is faint, try more cycles; too many cycles will results in band ‘spreading’. The best reassurance that correct genotypes are being obtained comes from the genotypes of standard individuals (such as the parents of CEPH pedigrees); failing that, Mendelian inheritance within established pedigrees can be checked.

Discussion The amount of labelled primer used in each reaction is a compromise that maximizes the efficient use of the radiolabel: using too little primer lowers the primer concentration and thus the efficiency of the PCR, while labelling too much primer (with the same amount of [γ - ^{32}P]ATP) results in low primer specific activity, and thus poor signal from those PCR products into which labelled primer was incorporated. In our experience, ^{33}P is a superior (but currently more expensive) alternative to ^{32}P for labelling primers. The main advantage of ^{33}P is its lower energy, so that it is safer to handle, and sharper bands are obtained on autoradiography.

In addition to end-labelled primers, other methods are can be used for typing dinucleotide repeats. PCR products can be detected after the incorporation of labelled nucleotides during amplification (e.g. ref. 134). The disadvantage of this method is that both product strands will be labelled. As the two strands will have different base compositions, they will have different mobilities on a polyacrylamide gel, and so two bands per allele will be detected [90]. While this should not often cause confusion, detecting one band per allele makes interpretation simpler. In the large-scale genetic typing of dinucleotide loci in the Généthon study [12], high throughput was achieved by doing multiplex PCR, blotting the gel onto a membrane and probing sequentially with locus- and strand-specific oligonucleotides. Fluorescent primers can also be used to type loci in conjunction with automated sequencing apparatus (see ref. 135).

.....

Troubleshooting

No signal/faint signal

This usually reflects poor overall PCR efficiency rather than poor primer labelling.

- Increase cycle number (but watch for band ‘spreading’)
- Check general PCR conditions (can use non-radioactive PCR and check on agarose gel)

- *If suspected, check incorporation of radioactivity into primer by running on a polyacrylamide gel.*

Too many bands: band ‘spreading’

A particular problem with dinucleotides is the tendency of PCR products to diversify during amplification. This problem becomes more pronounced with more PCR cycles.

- *Reduce the cycle number; use the smallest number of cycles that gives enough signal.*

Too many bands: additional ‘constant’ bands

Non-specific priming will give PCR products arising from other loci. This is a particular problem if the primer has some similarity to a dispersed repeat sequence.

- *Increase specificity of PCR conditions (annealing temperature, Mg²⁺ concentration; again, non-radioactive PCR and agarose gels can be used to establish these basic parameters).*
 - *If possible, label the other PCR primer instead.*
-

Acknowledgements

I am grateful for the support of the Wellcome Trust in my research: this chapter was mostly prepared during the tenure of a Wellcome Trust Postdoctoral Fellowship.

I am also very grateful to the following current and former colleagues who have contributed, in some cases unwittingly, to the material above: Alec Jeffreys, David Neil, Darren Monckton, Moira Crosier, Esther Signer, Nicola Royle, Zilla Wong, Rita Neumann, Katrina Mackay, Ila Patel, Raymond Dalgleish, Mark Jobling, Ian Gray, Vicky Wilson, Annette MacLeod, Maxine Allen, Stephanie Gobert, Duncan Baird, Andy Collick and Robert Kelly. Special thanks are due to Esther Signer and David Neil for foolishly agreeing to read and provide sensible comments on an earlier, less coherent version of this chapter.

References

1 Vogelstein, B., Fearon, E.R., Kern, S.E. *et al.* (1989) Allelo-type of colorectal carcinomas. *Science* **244**, 207–211.

2 Monckton, D. & Jeffreys, A.J. (1991) Minisatellite ‘isoallele’ discrimination in pseudohomozygotes by single molecule PCR and variant repeat mapping. *Genomics* **11**, 465–467.

3 Botstein, D., White, R.L., Skolnick, M. & Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am. J. Hum. Genet.* **32**, 314–331.

4 Reeders, S.T., Breuning, M.H., Davies, K.E. *et al.* (1985) A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* **317**, 542–544.

5 Steinlein, O., Fischer, C., Keil, R. *et al.* (1992) D20S19, linked to low voltage EEG, benign neonatal con-vulsions, and Fanconi anaemia, maps to a region of enhanced recombination and is localized between CpG islands. *Hum. Mol. Genet.* **1**, 325–329.

6 Chamberlain, S., Shaw, J., Rowland, A. *et al.* (1988) Mapping of mutation causing Friedrich’s ataxia to human chromosome 9. *Nature* **334**, 248–250.

7 Rouyer, F., Simmler, M.-C., Johnsson, C. *et al.* (1986) A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. *Nature* **319**, 291–295.

8 Petit, C., Levilliers, J. & Weissenbach, J. (1988) Physical mapping of the human pseudo-autosomal region; comparison with the genetic linkage map. *EMBO J.* **7**, 2369–2376.

9 Freije, D., Helms, C., Watson, M.S. & Donis-Keller, H. (1992) Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science* **258**, 1784–1787.

10 Gyapay, G., Morissette, J., Vignal, A. *et al.* (1994) The 1993–94 G  n  thon human genetic linkage map. *Nature Genet.* **7**, 246–339.

- 11 Reed, P.W., Davies, J.L., Copeman, J.B. *et al.* (1994) Chromosome-specific microsatellite sets for fluorescence-based, semiautomated genome mapping. *Nature Genet.* **7**, 390–395.
- 12 Weissenbach, J., Gyapay, G., Dib, C. *et al.* (1992) A second-generation linkage map of the human genome. *Nature* **359**, 794–801.
- 13 Dib, C., Faure, S., Fizames, C. *et al.* (1996) A comprehensive genetic map of the human genome based on 5264 microsatellites. *Nature* **380**, 152–154.
- 14 Bray-Ward, P., Menninger, J., Lieman, J. *et al.* (1996) Integration of the cytogenetic, genetic, and physical maps of the human genome by FISH mapping of CEPH YAC clones. *Genomics* **32**, 1–14.
- 15 Gyapay, G., Schmitt, K., Fizames, C. *et al.* (1996) A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5**, 339–346.
- 16 Hudson, T.J., Stein, L.D., Gerety, S.S. *et al.* (1995) An STS-based map of the human genome. *Science* **270**, 1945–54.
- 17 Chumakov, I.M., Rigault, P., Legall, I. *et al.* (1995) A YAC contig map of the human genome. *Nature* **377**, 175–297.
- 18 Buetow, K.H., Weber, J.L., Ludwigsen, S. *et al.* (1994) Integrated human genome-wide maps constructed using the CEPH reference panel. *Nature Genet.* **6**, 391–393.
- 19 Adamson, D., Albertsen, H., Ballard, L. *et al.* (1995) A collection of ordered tetranucleotide repeat markers from the human genome. *Am. J. Hum. Genet.* **57**, 619–628.
- 20 Matisse, T.C., Perlin, M. & Chakravarti, A. (1994) Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nature Genet.* **6**, 384–390.
- 21 Murray, J.C., Buetow, K.H., Weber, J.L. *et al.* (1994) A comprehensive linkage map with centimorgan density. *Science* **265**, 2049–2054.
- 22 White, R.L., Lalouel, J.-M., Nakamura, Y. *et al.* (1990) The CEPH consortium primary linkage map of human chromosome 10. *Genomics* **6**, 393–412.
- 23 Dracopoli, N., O'Connell, P., Elsner, T.I. *et al.* (1991) The CEPH consortium linkage map of human chromosome 1. *Genomics* **9**, 686–700.
- 24 Bowcock, A.M., Barnes, R.I., White, R.L. *et al.* (1992) The CEPH consortium linkage map of human chromosome 15q. *Genomics* **14**, 833–840.
- 25 Spurr, N.K., Cox, S., Bryant, S.P. *et al.* (1992) The CEPH consortium linkage map of human chromosome 2. *Genomics* **14**, 1055–1063.
- 26 Bowcock, A.M., Gerken, S.C., Barnes, R.I. *et al.* (1993) The CEPH consortium linkage map of human chromosome 13. *Genomics* **16**, 486–496.
- 27 Attwood, J., Chiano, M., Collins, A. *et al.* (1994) CEPH consortium map of chromosome 9. *Genomics* **19**, 203–214.
- 28 Kozman, H.M., Keith, T.P., Donis-Keller, H. *et al.* (1995) The CEPH consortium linkage map of human chromosome 16. *Genomics* **25**, 44–58.
- 29 Cox, D.W., Billingsley, G.D., Bale, A.E. *et al.* (1995) CEPH consortium map of chromosome-14. *Cytogenet. Cell Genet.* **69**, 175–178.
- 30 Litt, M., Kramer, P., Kort, E. *et al.* (1995) The CEPH consortium linkage map of human chromosome 11. *Genomics* **27**, 101–112.
- 31 Spurr, N.K., Bryant, S.P., Attwood, J. *et al.* (1994) European gene mapping project (EUROGEM): genetic maps based on the CEPH reference families. *Eur. J. Hum. Genet.* **2**, 193–252.
- 32 Fain, P.R., Kort, E.N., Chance, P.F. *et al.* (1995) A 2D crossover-based map of the human X-chromosome as a model for map integration. *Nature Genet.* **9**, 261–266.
- 33 Gerken, S.C., Albertsen, H., Elsner, T. *et al.* (1995) A strategy for constructing high-resolution genetic maps of the human genome—a genetic map of chromosome 17p, ordered with meiotic breakpoint mapping panels. *Am. J. Hum. Genet.* **56**, 484–499.
- 34 Donis-Keller, H., Green, P., Helms, C. *et al.* (1987) A genetic linkage map of the human genome. *Cell* **51**, 319–337.
- 35 O'Connell, P., Lathrop, G.M., Law, M. *et al.* (1987) A primary genetic linkage map for human chromosome 12. *Genomics* **1**, 93–102.
- 36 O'Connell, P., Lathrop, G.M., Leppert, M. *et al.* (1988) Twelve loci form a continuous linkage map for human chromosome 18. *Genomics* **3**, 367–372.
- 37 Nakamura, Y., Lathrop, M., O'Connell, P. *et al.* (1988) A primary map of 10 DNA markers and two serological markers for human chromosome 19. *Genomics* **3**, 67–71.
- 38 Nakamura, Y., Lathrop, M., O'Connell, P. *et al.* (1988) A mapped set of markers for human chromosome 17. *Genomics* **2**, 302–309.
- 39 Nakamura, Y., Lathrop, M., O'Connell, P. *et al.* (1989) Frequent recombination is observed in the distal end of the long arm of chromosome 14. *Genomics* **4**, 76–81.
- 40 NIH/CEPH Collaborative Mapping Group (1992) A comprehensive genetic linkage map of the human genome. *Science* **258**, 148–162.
- 41 Levitt, R.C. (1991) Polymorphisms in the transcribed 3' untranslated region of eukaryotic genes. *Genomics* **11**, 484–489.
- 42 Avramopoulos, D., Chakravarti, A. & Antonarakis, S.E. (1993) DNA polymorphisms in the 3' untranslated region of genes on human chromosome 21. *Genomics* **15**, 98–102.
- 43 Litt, M. & White, R.L. (1985) A highly polymorphic locus in human DNA revealed by cosmid-derived probes. *Proc. Natl Acad. Sci. USA* **82**, 6202–6210.
- 44 Nakamura, Y., Leppert, M., O'Connell, P. *et al.* (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**, 1616–1622.
- 45 Nakamura, Y., Carlson, M., Krapcho, K. *et al.* (1988) New approach for isolation of VNTR markers. *Am. J. Hum. Genet.* **43**, 854–859.
- 46 Schumm, J., Knowlton, R., Braman, J. *et al.* (1985) Detection of more than 500 single-copy RFLPs by random screening. *Cytogenet. Cell Genet.* **40**, 739.
- 47 Knowlton, R., Schumm, J., Braman, J. *et al.* (1985) Relative efficiencies of restriction endonucleases in detecting RFLPs. *Cytogenet. Cell Genet.* **40**, 670.
- 48 Hiyashi, K. (1991) PCR-SSCP: a simple and sensitive method for the detection of mutations in genomic DNA. *PCR Meth. Appl.* **1**, 34–38.

- 49 Sheffield, V., Cox, D.R., Lerman, L.S. & Myers, R.M. (1989) Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic fragments by the polymerase chain reaction results in improved detection of single-base-changes. *Proc. Natl Acad. Sci. USA* **86**, 232–236.
- 50 Riesner, D., Steger, G., Zimmat, R. *et al.* (1989) Temperature-gradient gel-electrophoresis of nucleic acids-analysis of conformational transitions, sequence variations, and protein–nucleic acid interactions. *Electrophoresis* **10**, 377–389.
- 51 White, M.B., Carvalho, M., Derse, D. *et al.* (1992) Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* **12**, 301–306.
- 52 Cotton, R.G.H., Rodrigues, N.R. & Campbell, D.R. (1988) Reactivity of cytosine and thymine in single base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Natl Acad. Sci. USA* **85**, 4397–4401.
- 53 Myers, R.M., Larin, Z. & Maniatis, T. (1985) Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* **230**, 1242–1246.
- 54 Underhill, P.A., Jin, L., Zeman, R. *et al.* (1996) A preColumbian Y-chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl Acad. Sci. USA* **93**, 196–200.
- 55 Youil, R., Kemper, B.W. & Cotton, R.G.H. (1995) Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. *Proc. Natl Acad. Sci. USA* **92**, 87–91.
- 56 Orita, M., Suzuki, Y., Sekiya, T. & Hiyashi, K. (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874–879.
- 57 Iwahana, H., Yoshimoto, K. & Itakura, M. (1992) Detection of point mutations by SSCP of PCR-amplified DNA after endonuclease digestion. *BioTechniques* **12**, 64.
- 58 Mackay, K., Byers, P.H. & Dalgleish, R. (1993) An RT-PCR-SSCP screening strategy for detection of mutants in the gene encoding the $\alpha 1$ chain of type I collagen: application to four patients with osteogenesis imperfecta. *Hum. Mol. Genet.* **2**, 1155–1160.
- 59 Armour, J.A.L., Harris, P.C. & Jeffreys, A.J. (1993) Allelic diversity at minisatellite MS205 (D16S309): evidence for polarized variability. *Hum. Mol. Genet.* **2**, 1137–1145.
- 60 Nickerson, D.A., Whitehurst, C., Boysen, C. *et al.* (1992) Identification of clusters of biallelic polymorphic sequence-tagged sites (pSTSs) that generate highly informative and automatable markers for genetic linkage mapping. *Genomics* **12**, 377–387.
- 61 Bevan, I.S., Rapley, R. & Walker, M.R. (1992) Sequencing of PCR-amplified DNA. *PCR Meth. Appl.* **2**, 222–228.
- 62 Winship, P. (1989) An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Res.* **17**, 1266.
- 63 Casanova, J.-L., Pannetier, C., Jaulin, C. & Kourilsky, P. (1990) Optimal conditions for directly sequencing double-stranded PCR products with Sequenase. *Nucleic Acids Res.* **18**, 4028.
- 64 Murray, V. (1989) Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res.* **17**, 8889.
- 65 Srivastava, A.K., Montanaro, V. & Kere, J. (1992) Simplified template preparation and improved direct sequencing using *Taq* polymerase. *PCR Meth. Appl.* **2**, 255–256.
- 66 Gyllenstein, U.B. & Erlich, H.A. (1988) Generation of single-stranded DNA by polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc. Natl Acad. Sci. USA* **85**, 7652–7656.
- 67 Higuchi, R.G. & Ochman, H. (1989) Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. *Nucleic Acids Res.* **17**, 5865.
- 68 Hultman, T., Stahl, S., Hornes, E. & Uhlen, M. (1989) Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. *Nucleic Acids Res.* **17**, 4937–4946.
- 69 Sarkar, B. & Sommer, S. (1988) RNA amplification with transcript sequencing (RAWTS) *Nucleic Acids Res.* **16**, 5197.
- 70 Marchuk, D., Drumm, M., Saulino, A. & Collins, F.S. (1991) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* **19**, 1154.
- 71 Kovalic, D., Kwak, J.-H. & Weisblum, B. (1991) General method for direct cloning of DNA fragments generated by the polymerase chain reaction. *Nucleic Acids Res.* **19**, 4560.
- 72 Crowe, J.S., Cooper, H.J., Smith, M.A. *et al.* (1991) Improved cloning efficiency of polymerase chain reaction (PCR) products after proteinase K digestion. *Nucleic Acids Res.* **19**, 184.
- 73 Li, H., Gyllenstein, U.B., Cui, X. *et al.* (1988) Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* **335**, 414–417.
- 74 Nickerson, D.A., Kaiser, R., Lappin, S. *et al.* (1990) Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay. *Proc. Natl Acad. Sci. USA* **87**, 8923–8927.
- 75 Syvanen, A.C., Sajantila, A. & Lukka, M. (1993) Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing. *Am. J. Hum. Genet.* **52**, 46–59.
- 76 Syvanen, A.C. & Landegren, U. (1994) Detection of point mutations by solid-phase methods. *Hum. Mutat.* **3**, 172–179.
- 77 Newton, C.R., Graham, A., Heptinstall, I.E. *et al.* (1989) Analysis of any point mutation in DNA: the amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17**, 2503–2515.
- 78 Sommer, S.S., Groszback, A.R. & Bottema, C.D.K. (1992) PCR amplification of specific alleles (PASA) is a general method for rapidly detecting known single-base changes. *BioTechniques* **12**, 82–87.
- 79 Wu, D.Y., Ugozzoli, L., Pal, B.K. & Wallace, R.B. (1989) Allele-specific enzymatic amplification of β -globin genomic DNA for diagnosis of sickle cell anaemia. *Proc. Natl Acad. Sci. USA* **86**, 2757–2760.

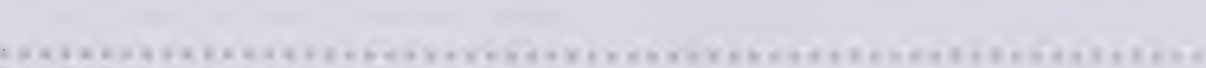
- 80 Neil, D.L. & Jeffreys, A.J. (1993) Digital DNA typing at a second hypervariable locus by minisatellite variant repeat mapping. *Hum. Mol. Genet.* **2**, 1129–1135.
- 81 Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
- 82 Kwiatkowski, D.J., Henske, E.P., Weimer, K. *et al.* (1992) Construction of a GT polymorphism map of human 9q. *Genomics* **12**, 229–240.
- 83 Schmid, C.W. & Jelinek, W.R. (1982) The Alu family of dispersed repetitive sequences. *Science* **216**, 1065–1070.
- 84 Telenius, H., Carter, N.P., Bebb, C.E. *et al.* (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* **13**, 718–725.
- 85 Grothues, D., Cantor, C.R. & Smith, C.L. (1993) PCR amplification of megabase DNA with tagged random primers (T-PCR). *Nucleic Acids Res.* **21**, 1321.
- 86 Armour, J.A.L., Povey, S., Jeremiah, S. & Jeffreys, A.J. (1990) Systematic cloning of human minisatellites from ordered array Charomid libraries. *Genomics* **8**, 501–512.
- 87 Weber, J.L. (1990) Informativeness of human (dA-dC)n. (dG-dT)n polymorphisms. *Genomics* **7**, 524–530.
- 88 Litt, M. & Luty, J.A. (1989) A hypervariable microsatellite revealed by in vitro amplification of a nucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44**, 397–401.
- 89 Weber, J.L. & May, P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388–396.
- 90 Tautz, D. (1989) Hypervariability of simple sequences as a general source for polymorphic markers. *Nucleic Acids Res.* **17**, 6463–6471.
- 91 Fu, Y.-H., Pizzuti, A., Fenwick, R.G. *et al.* (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**, 1256–1259.
- 92 Rychlik, W. & Rhoads, R.E. (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Res.* **17**, 8543–8551.
- 93 Hudson, T.J., Engelstein, M., Lee, M.K. *et al.* (1992) Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. *Genomics* **13**, 622–629.
- 94 Jeffreys, A.J., Wilson, V., Thein, S.L. *et al.* (1986) DNA 'fingerprints' and segregation analysis of multiple markers in human pedigrees. *Am. J. Hum. Genet.* **39**, 11–24.
- 95 Armour, J.A.L., Vergnaud, G., Crosier, M. & Jeffreys, A.J. (1992) Isolation of human minisatellite loci detected by synthetic tandem repeat probes: direct comparison with cloned DNA fingerprinting probes. *Hum. Mol. Genet.* **1**, 319–323.
- 96 Armour, J.A.L. & Jeffreys, A.J. (1992) Biology and applications of human minisatellite loci. *Curr. Opin. Genet. Dev.* **2**, 850–856.
- 97 Spurr, N.K. (1991) Standards for reporting alleles at highly polymorphic loci: a proposal. *Hum. Mol. Genet.* **1**, 211–212.
- 98 Loupart, M.-L., Armour, J.A.L., Jeffreys, A.J. & Varley, J.M. (1991) 33.6 (D1S111) and pMLA1 (D1S61) identify the same VNTR on chromosome 1. *Nucleic Acids Res.* **19**, 4801.
- 99 Economou, E.P., Bergen, A.W., Warren, A.C. & Antonarakis, S.E. (1990) The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc. Natl Acad. Sci. USA* **87**, 2951–2954.
- 100 Kalaitzidaki, M., Cox, T., Chakravarti, A. & Antonarakis, S.E. (1992) Cloning and linkage mapping of three polymorphic tetranucleotide (TAAA)_n repeats on human chromosome 21. *Genomics* **14**, 1071–1075.
- 101 Singer, M.F. & Skowronski, J. (1985) Making sense out of LINES: long interspersed repeat sequences in mammalian genomes. *Trends Biochem. Sci.* **10**, 119–122.
- 102 Lagerstrom, M., Parik, J., Malmgren, H. *et al.* (1991) Capture PCR: efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA. *PCR Meth. Appl.* **1**, 111–119.
- 103 Karagyasov, L., Kalcheva, I.D. & Chapman, V.M. (1993) Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. *Nucleic Acids Res.* **21**, 3911–3912.
- 104 Kandpal, R.P., Kandpal, G. & Weissman, S.M. (1994) Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. *Proc. Natl Acad. Sci. USA* **91**, 88–92.
- 105 Yuille, M.A.R., Goudie, D.R., Affara, N.A. & Ferguson-Smith, M.A. (1991) Rapid determination of sequences flanking microsatellites. *Nucleic Acids Res.* **19**, 1950.
- 106 Edwards, A., Civitello, A., Hammond, H.A. & Caskey, C.T. (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* **49**, 746–756.
- 107 Caskey, C.T., Pizzuti, A., Fu, Y.-H. *et al.* (1992) Triplet mutations in human disease. *Science* **256**, 784–789.
- 108 Huntington's Disease Collaborative Group. (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983.
- 109 Riggins, G.J., Lokey, L.K., Chastain, J.L. *et al.* (1992) Human genes containing polymorphic trinucleotide repeats. *Nature Genet.* **2**, 186–191.
- 110 Li, S.-H., McInnis, M.G., Margolis, R.L. *et al.* (1993) Novel triplet repeat containing genes in human brain: cloning, expression, and length polymorphisms. *Genomics* **16**, 572–579.
- 111 Melis, R., Bradley, P., Elsner, T. *et al.* (1993) Polymorphic SSR (simple sequence repeat) markers for chromosome 20. *Genomics* **16**, 56–62.
- 112 Armour, J.A.L., Neumann, R., Gobert, S. & Jeffreys, A.J. (1994) Isolation of human simple repeat loci by hybridization selection. *Hum. Mol. Genet.* **3**, 599–605.
- 113 Boerwinkle, E., Xiong, W., Fourest, E. & Chan, L. (1989) Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction: application to the apolipoprotein B 3' hypervariable region. *Proc. Natl Acad. Sci. USA* **86**, 212–216.

- 114 Budowle, B., Chakraborty, R., Giusti, A.M. *et al.* (1991) Analysis of the VNTR locus D1S80 by PCR followed by high-resolution PAGE. *Am. J. Hum. Genet.* **48**, 137–144.
- 115 Jeffreys, A.J., Wilson, V., Neumann, R. & Keyte, J. (1988) Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Res.* **16**, 10953–10971.
- 116 Jeffreys, A.J., Royle, N.J., Wilson, V. & Wong, Z. (1988) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* **332**, 278–281.
- 117 Vergnaud, G., Mariat, D., Apiou, F. *et al.* (1991) The use of synthetic tandem repeats to isolate new VNTR loci: cloning of a human hypermutable sequence. *Genomics* **11**, 135–144.
- 118 Wong, Z., Wilson, V., Jeffreys, A.J. & Thein, S.L. (1986) Cloning a selected fragment from a human DNA 'fingerprint': isolation of an extremely polymorphic minisatellite. *Nucleic Acids Res.* **14**, 4605–4616.
- 119 Wong, Z., Wilson, V., Patel, I. *et al.* (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.* **51**, 269–288.
- 120 Zischler, H., Kammerbauer, C., Studer, R. *et al.* (1992) Dissecting (CAC)₅/ (GTG)₅ multilocus fingerprints from man into individual locus-specific, hypervariable components. *Genomics* **13**, 983–990.
- 121 Hanotte, O., Burke, T., Armour, J.A.L. & Jeffreys, A.J. (1991) Hypervariable minisatellite DNA sequences in the peafowl *Pavo cristatus*. *Genomics* **9**, 587–597.
- 122 Georges, M., Gunawardana, A., Threadgill, D. *et al.* (1991) Characterization of a set of variable number of tandem repeat markers conserved in bovidae. *Genomics* **11**, 24–32.
- 123 Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**, 67–73.
- 124 Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985) Individual-specific 'fingerprints' of human DNA. *Nature* **316**, 76–79.
- 125 Hill, A.V.S. & Jeffreys, A.J. (1985) Use of minisatellite DNA probes for determination of twin zygosity at birth. *Lancet* **ii** 1394–1395.
- 126 Welsh, J. & McClelland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* **18**, 7213–7218.
- 127 Williams, J.G.K., Kubelik, A.R., Livak, K.J. *et al.* (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**, 6531–6535.
- 128 Riedy, M.F., Hamilton, W.J. III & Aquadro, C.F. (1992) Excess of nonparental bands in offspring from known primate pedigrees assayed using RAPD PCR. *Nucleic Acids Res.* **20**, 918.
- 129 Nelson, S.F., McCusker, J.H., Sander, M.A. *et al.* (1993) Genomic mismatch scanning: a new approach to genetic linkage mapping. *Nature Genet.* **4**, 11–17.
- 130 Lisitsyn, N., Lisitsyn, N. & Wigler, M. (1993) Cloning the differences between two complex genomes. *Science* **259**, 946–951.
- 131 Schutte, M., Dacosta, L.T., Hahn, S.A. *et al.* (1995) Identification by representational difference analysis of a homozygous deletion in pancreatic carcinoma that lies within the BRCA2 region. *Proc. Natl Acad. Sci. USA* **92**, 5950–5954.
- 132 Vos, P., Hogers, R., Bleeker, M. *et al.* (1995) AFLP—a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414.
- 133 Feinberg, A.P. & Vogelstein, B. (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**, 266–267.
- 134 Coleman, M.P., Murray, J.C., Willard, H.F. *et al.* (1991) Genetic and physical mapping around the properdin P gene. *Genomics* **11**, 991–996.
- 135 Ziegler, J.S., Su, Y., Corcoran, K.P. *et al.* (1992) Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics* **14**, 1026–1031.
- 136 Bains, W. (1986) The multiple origins of human Alu elements. *J. Mol. Evol.* **23**, 189–199.

Chapter 6 DNA typing

John A.L. Armour

Department of Genetics, School of Medicine, Queen’s Medical Centre, Nottingham NG7 2UH, UK



6.1	Introduction, 128	
6.1.1	What DNA typing is for, 128	
6.1.2	Planning and general strategy: doing it the easy way, 128	
6.1.3	DNA typing systems, 128	
6.1.4	Statistical evaluation: general notes, 129	
6.1.4.1	Matches and probabilities: two simple cases, 129	
6.1.4.2	Allele frequencies and population substructuring, 130	
6.1.5	Disclaimer: learned friends, please note, 130	
6.2	DNA fingerprinting, 131	
6.2.1	General principles, 131	
6.2.2	DNA fingerprinting in practice, 132	
6.2.3	Data interpretation and statistical evaluation, 132	
6.3	Locus-specific minisatellite probes, 132	
6.3.1	Advantages and disadvantages, 132	
6.3.2	Typing individual minisatellite loci: data interpretation and statistical evaluation, 133	
6.3.2.1	Match criteria, 133	
6.3.2.2	Mutation or exclusion?, 133	
6.3.2.3	Allele frequencies and probabilities, 133	
6.4	PCR analysis of minisatellites, 134	
6.4.1	AMPFLPs and MVR-PCR: advantages and disadvantages, 134	
6.4.2	Available loci, 134	
6.4.3	Data interpretation and statistical evaluation, 135	
6.4.3.1	AMPFLPs, 135	
6.4.3.2	MVR-PCR, 135	
6.5	PCR typing using simple tandem repeats, 135	
6.5.1	Advantages and disadvantages, 135	
6.5.2	Typing STR loci, 135	
6.5.3	Data interpretation and statistical evaluation, 135	
	Protocol 10	
	DNA fingerprinting, 136	
	Troubleshooting, 138	
	No signal/faint signal, 138	
	Too many bands, 138	
	Protocol 11	
	Typing using locus-specific minisatellite probes, 138	
	Troubleshooting, 139	
	Not enough signal, 139	
	Too much signal/complex profile, 140	
	Protocol 12	
	MVR-PCR at D1S8, 140	
	Troubleshooting, 141	
	Not enough signal, 141	
	Most signal in smallest fragments, 142	
	Acknowledgements, 142	
	References, 142	

6.1 Introduction

6.1.1 What DNA typing is for

DNA typing uses the information from DNA polymorphisms to resolve issues of biological origin, most frequently concerning identity or genetic relationship. As in more general uses of DNA polymorphisms in research, the most powerful systems are those with the most extensive variation. The best-known applications for DNA typing (see Applications box 6.1) are two situations important in family law—paternity testing [1–3] and the verification of family relationships in immigration [4]—and in forensic cases in establishing a link between DNA in (for example) blood or semen found at a crime scene and a suspect’s DNA [5]. DNA typing, however, also has important applications in genetic research, such as quality control of DNA samples or cell-line identity [6] and in the establishing of twin zygoty [7]. This chapter is intended to be an introduction to the practical implementation of DNA typing in the context of genetic research (see also Section 1.5).

A note on nomenclature The term ‘DNA fingerprinting’ will be used in this chapter in its original sense [2], to refer only to typing using DNA profiles composed of contributions from large numbers of hypervariable loci, such that the combined pattern will be *individual-specific* (hence ‘fingerprint’). The term ‘DNA fingerprinting’ has, however, since acquired a wider meaning, applying to a wide variety of DNA-based profiling methods, many of which fall well short of individual specificity.

6.1.2 Planning and general strategy: doing it the easy way

Each of the general systems for DNA typing outlined in Section 6.1.3 have their own advantages and disadvantages in practice, and the right system

for the job in hand will depend on its resolving power for the application required, and also its practical convenience, both in terms of technical simplicity and the amount of work needed for a test. The relative merits of the different systems discussed will be reviewed overall in Section 6.1.3 and in detail in Sections 6.2.1, 6.3.1, 6.4.1 and 6.5.1.

It is worth remembering, however, that in very many cases the samples in question will have been the subject of a number of genetic analyses as part of the central study. These data will themselves allow simple deductions about biological identity. For example, segregation analysis in a pedigree using five or six dinucleotide repeat loci may reveal enough discrepancies to cast doubt on paternity, or to raise suspicion of a sample mix-up. Furthermore, standard estimates of allele frequencies will be readily available for commonly used dinucleotide repeat loci, allowing simple evaluation of statistical issues (see also Section 6.5.3).

Two systems that will not be discussed in detail but merit at least an introductory mention, are DNA-based HLA-typing and the use of mitochondrial DNA. DNA-based typing of HLA variation, and in particular the HLA-DQ α locus, has the advantage of simplicity and a PCR-based assay—indeed a kit for DNA typing of DQ α is commercially available (Perkin-Elmer). Its utility is limited by the low variability of the locus [8], as well as the fact that only a single locus is involved (and therefore it has poor resolving power in distinguishing sibs). Mitochondrial DNA typing, whether by sequence analysis or restriction mapping, has been applied with great success to highly degraded DNA samples in one high-profile study [9]. Its resolving power, however, is relatively poor, being unable to distinguish even quite distant maternal relatives, such as HRH the Duke of Edinburgh and Tsarina Alexandra [9].

6.1.3 DNA typing systems

The different systems available for high-resolution DNA typing discussed here are summarized in Table 6.1. For sheer statistical power, DNA fingerprinting remains the single best technique, since it combines information from many highly informative loci in a single test. It is very powerful in resolving all types of issues in DNA typing, and the chief disadvantage is that relatively large amounts (> 1 μ g per test) of high-quality DNA are required for good results. Researchers trying the technique for the first time may also find that results improve with practice.

Typing minisatellite loci singly, using single-locus

DNA typing in humans is used in:

- parentage and family analysis
- individual identification, especially in forensic science

Important research applications include:

- checking family relationships (especially paternity) in pedigrees
- checking identity, for example of DNA samples or cell lines
- checking twin zygoty

Applications box 6.1

Table 6.1 Properties of DNA typing systems.

	Discrimination power ^a			Amount of DNA needed	Technical simplicity ^b
	Parentage	Unrelated	Siblings		
DNA fingerprinting	+++	+++	+++	Large	+/-
Single locus minisatellites	++	++	++	Moderate	++
PCR typing minisatellites	++	++	++	Small	+/-
MVR-PCR	+/-	+++	-	Small	+/+
Simple tandem repeats	+	+	+	Small	+++

^a ‘Discrimination power’ summarizes the relative resolving power *per experiment*.
^b ‘Technical simplicity’ is to some extent operator dependent, and this is reflected in the ambiguities in this column.

probes, gives less information per experiment, but combining results from a number of different loci can give high cumulative resolving power. Smaller amounts of DNA (0.1–1 µg per test) can be used than are required for good DNA fingerprints. Although they consume relatively large amounts of high-quality DNA, one particular advantage of tests based on Southern blot hybridization is that the same Southern blots can be used for many probes, stripping off each probe before re-use. A set of filters, for example, might be used for two DNA fingerprinting probes and then for a number of single-locus profiles. Single-locus profiles or DNA fingerprinting may present a technical barrier in PCR-monopolized laboratories.

PCR analysis of minisatellites and minisatellite variant repeat PCR (MVR-PCR) combine some of the resolving power of minisatellite loci with the sensitivity of PCR. The disadvantages of PCR typing of minisatellites are that relatively few well-characterized systems are available, and care needs to be taken to make sure that relatively large alleles are not selected against during PCR. MVR-PCR can produce extremely informative profiles from small amounts of DNA, and its digital results give simple and objective match criteria. It has the disadvantage that only one locus has so far given simple codes from diploid DNA, and thus the system has limited resolving power in paternity analysis and in distinguishing siblings.

PCR analysis of simple tandem-repeat loci has the advantage that numerous well-documented systems are available and the technology for typing them will be familiar to most laboratories. Indeed, as outlined in Section 6.1.2, typing such loci will frequently be an intrinsic part of the overall study. The main disadvantage is that each system will have limited resolving power, such that satisfactory statistical weight can only be derived from typing a large number of loci.

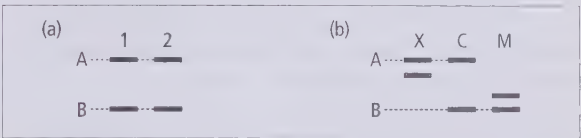


Fig. 6.1 Schematic representation of DNA typing results from a single polymorphic locus in the analysis of (a) identity between samples 1 and 2, and (b) paternity. X, alleged father; C, child; M (undisputed) mother. A and B are the two alleles at that locus.

6.1.4 Statistical evaluation: general notes

The interpretation of data and the statistical evaluation of results will be discussed in detail in the individual sections relating to each method of DNA typing. However, some more general points, applying to DNA typing in general, will be introduced here. In general, the main question will reduce to ‘how much more (or less) likely are these DNA typing results under hypothesis A compared with hypothesis B?’. This in turn will generally be deduced from known allele or band frequencies deduced from previous work with these systems. The first point to make is that the power required of a test depends very much on the question being asked. Thus, ‘are these two samples from the same person, or has there been a sample mix-up?’ will require relatively low-resolution information. More detailed questions, such as ‘is this man or his brother the father of this child?’ require more detailed information. The basic considerations surrounding the commonest questions are illustrated for a single (biparentally inherited) locus test in Fig. 6.1a (a test of identity: identical twins/sample mix-up) and Fig. 6.1b (a test of paternity).

6.1.4.1 Matches and probabilities: two simple cases

In Fig. 6.1a, a ‘positive’ result is obtained when the two samples match at both alleles of this locus; how

unlikely is this result if the two samples are in fact from different individuals? This depends on the allele frequencies of the two alleles A and B . If the test is of identity vs. a random comparison in the population, then the probability of this result occurring by chance alone is twice the product of the two allele frequencies, i.e. $2q_A \times q_B$. If, however, the test is of twin zygosity—in other words, of identity vs. sibship—then the significance will depend on the parental genotypes; if one parent is homozygous for allele A and the other for allele B at this locus, then in the absence of mutation this will be the only possible result for any of their offspring, and thus the test has no resolving power at all. By contrast, if the parents have between them four different alleles at this locus, of which A and B are two, then the probability that two sibs will happen to share a genotype (AB) will be 0.25.

Paternity analysis relies on the fact that in the absence of mutation, one of a child's alleles at each locus will be maternally inherited, the other paternally. In Fig. 6.1b, the child inherits allele B from the (undisputed) mother, and its other allele (allele A) must have come from the true father. Individual X has an allele A which thus is *either* the allele he passed on to his offspring *or* an allele which just happens to be of the right size. How much of a coincidence it is that he just happens to have a band of the right size will depend on the population frequency of that allele (q_A), and will be given by $2q_A - q_A^2$ (either one of his two alleles could be A , giving $2q_A(1 - q_A)$, or he could be homozygous for A , giving q_A^2 ; the sum of these is $2q_A - q_A^2$).

The examples given above simply illustrate some of the basic principles used in calculating probabilities, and have neglected circumstances which make the reasoning more involved (such as when an allele is shared by mother and alleged father, or where the true father and alleged father may be close relatives), or complications such as germline mutation which are important in practice when using the relatively unstable (but highly informative) minisatellite loci.

6.1.4.2 Allele frequencies and population substructuring

It has so far been assumed that the allele frequencies are known with some accuracy for the typing system in question, so that correct probabilities can be calculated in a straightforward way. For many systems these values will have been derived from extensive empirical observation. A general point to remember, though, is that most allele frequency values have been determined from outbred (and generally European) 'reference' populations. If the test applies to individuals from a very different

population, then the frequencies of a particular allele may be higher for all members of that population; since this effect may apply at more than one locus, for the evaluation of probabilities the different loci cannot fairly be regarded as independent, unless frequency data are available for that population group.

It is a matter of common observation that humans worldwide do not form one large outbred population, but mate assortatively in smaller groups; this may therefore mean that alleles which are rare in the reference population may fortuitously have a higher frequency in the population group under study, thereby diminishing the significance of a match. To take one extreme case, it would be inappropriate to apply 'standard' European allele frequencies uncritically to analyses of individuals from Eastern Turkey, not only because the alleles in question may have different frequencies in that relatively isolated population, but also because they are more highly inbred (more than 20% of marriages in Eastern Turkey are between first cousins [10]) and thus a particular allele may have an unusually high frequency in an extended group of local families.

Few doubt that there are indeed subpopulations within *Homo sapiens*, but few also would consider the magnitude of these effects to be significant in large cosmopolitan populations given the high variability of most polymorphisms used in DNA typing. The extent to which loci can be regarded as independent is an important issue in the application of DNA typing in legal contexts, and has been the subject of extensive (and not always good-tempered) debate [11–19]. One solution proposed has been to assemble allele frequency databases for each population, but this in turn raises the very difficult question of what constitutes a distinct human 'population'.

Nevertheless, since this discussion of DNA typing is intended to give guidance on its use in a research context, and thus to not to be concerned with the minutiae of the arguments (see Section 6.1.5), statistical evaluation will proceed on the assumption that fair estimates of allele frequencies are known for the samples in question. It should be noted that two systems, DNA fingerprinting (Section 6.2) and MVR-PCR (Section 6.4) do not score *alleles*. The empirical methods used in assessing their power of resolution need not depend on population homogeneity (see Sections 6.2.3 and 6.4.3).

6.1.5 Disclaimer: learned friends, please note

Most of us—excepting only the most stringent

experimentalists—would agree that higher standards of evidence and a greater burden of proof are required of DNA typing in a legal context than of DNA typing in research applications. Although high standards should apply in the use of DNA typing in genetic research, unusually robust standards must clearly apply where the loss of liberty or attribution of paternity are at stake. It is therefore important to make absolutely clear that the advice offered in this chapter is intended as a guide to the use of DNA typing in the context of genetic research, and *not* in situations where the results may carry legal weight. This consideration particularly applies to the determination of statistical weight in the evaluation of results; the intention in the discussions below is to use reasonable approximations and assumptions to allow very probable inferences to be drawn, rather than to robustly exclude even the most unusual coincidences.

Those using DNA typing of humans exclusively for research purposes should also note that even routine genotyping could (indirectly) provide unexpected information about pedigrees under study, especially unexpected exclusions of paternity. While this would have important implications for linkage analysis studies, it could clearly also have important legal and psychological implications for the family if it became known; remember that in most cases the family will have consented to the study of their inherited disorder, and will *not* have asked for paternity testing. These considerations should emphasize the importance of maintaining absolute confidentiality in studies of human pedigrees, and in particular the benefits of *anonymous coding* of sample identities before they reach the DNA laboratory. It should also be obvious to the thoughtful reader why it is a very dangerous practice to use laboratory workers and their (assumed) kin as ‘control families’ in genetic analysis.

6.2 DNA fingerprinting

6.2.1 General principles

The basic principle underlying multilocus DNA fingerprinting is that many tandemly repetitive probes will cross-hybridize with other, non-cognate loci under conditions of low stringency. The most useful probes are those which cross-hybridize with a large number of other loci, many of which are themselves highly polymorphic. The combined resolving power of analysing many hypervariable loci in a single profile (the ‘DNA fingerprint’) makes the result individual-specific [1,2,20]—with the obvious exception of identical twins [7]. Many

probes have now been shown to detect multiple hypervariable loci; some of the more widely used are listed in Table 6.2.

The main advantage of using multilocus DNA fingerprinting is that a large number of loci can be scanned simultaneously; this gives great resolving power in the analysis of parentage, individual identity and family relationships [1,2,20] and in analysing loss of heterozygosity in tumours (J.A.L. Armour, unpublished). Although of great analytical value in these contexts, and of particular application in distinguishing close relatives, its value in linkage analysis is much less. The main disadvantages are: firstly, that although unlinked, the loci detected are not distributed randomly in the genome, but have a strong tendency to localize to subtelomeric regions [21,22]; secondly, usually only one of a pair of alleles in a heterozygote can be resolved, with the result that linkage analysis is only appropriate for autosomal dominant traits [20]; thirdly (and most serious practically), profiles from different families can not be directly compared, as bands in similar locations on the profile in different families will almost certainly derive from different loci—the whole point is that most of the loci scored are extremely variable in size; fourthly, converting from an interesting band on a DNA fingerprint to a corresponding locus-specific probe is technically laborious [23].

Thus, in summary, the main uses for multilocus DNA fingerprinting are: (a) in purely analytical situations, such as parentage verification or zygosity testing in twins, where the identity and location of the loci involved are not at issue; (b) in linkage analysis with autosomal dominant traits (or *between* fingerprint bands) within a single kindred, or when the definition of allelic series is unambiguous, for example in a single large kindred [20,24], or with reference to inbred founder organisms [25,26].

Table 6.2 Multilocus DNA fingerprinting probes.

Probe	Type ^a	Ref.
33.6	a	[1,2,20]
33.15	a	[1,2,20]
(CAC) <i>n</i>	c	[21]
M13	b	[22]
‘STR’	d	[23]
α-globin 3’HVR	a	[24,25]
YNZ2	a	[26]

^aProbe types: a, human genomic DNA clone; b, phage genomic repeat region; c, synthetic oligonucleotide; d, synthetic tandem repeat array.

6.2.2 DNA fingerprinting in practice

Protocol 10 outlines our own approach to producing high-quality, informative DNA fingerprints using probes 33.6 and 33.15 [1,2]. The details of probe preparation are not given here, simply because of the diversity of methods available.

6.2.3 Data interpretation and statistical evaluation

DNA fingerprinting profiles give a set of bands for each individual composed of contributions from a large number of different loci. The well-resolved fragments in the larger size range (> 2 kb) give most information, and the total number of bands scored per person in this size range will obviously vary. The statistical evaluation of the results will be discussed with reference to a large data set based on fragments > 3.5 kb [3]. Note, however, that although this size range has been chosen for statistical analysis, qualitative information will also be available from other parts of the profile: patterns matching or discordant in the > 3.5 kb range will also show similar results among smaller bands.

Extensive casework has shown that a value of 0.25 is a conservative estimate of the average frequency for band-sharing > 3.5 kb between unrelated individuals [3]. Technical issues of what constitutes a match between bands in different lanes has been a long-standing source of interest to critics of the legal uses of DNA typing, and will be discussed further in Section 6.3.2. The simple (and conservative) criterion used by Jeffreys *et al.* [3] assumes that the true positions of two bands of indistinguishable mobility may actually differ by as much as 0.5 mm.

In a comparison of two samples for identity, the probability that n bands, at a probability of 0.25 per band, will be shared purely by chance between unrelated individuals is $(0.25)^n$, assuming band independence. For paternity evaluation, we are concerned with the number (x) of *non-maternal* bands in the child which are shared by the supposed father; if these are all present in the supposed father,

then the probability of this occurring purely by chance is $(0.25)^x$. If many of these non-maternal bands (more than 40%) are missing from the father, paternity is excluded. The complication of using highly polymorphic (and therefore highly unstable) minisatellite loci is that germline mutation will occasionally happen. Thus, a band may be present in the child which is not found in the mother or the father even though they are the real parents of that child. In practice, empirical observation has shown that if the parentage is correct, fewer than 20% of the non-maternal bands in the child will be ‘unassignable’ if they are due to real germline mutations (rather than incorrect paternity). The guidelines given above are all based on conservative extrapolations from a large data set of paternity casework: for detailed discussion see [3].

6.3 Locus-specific minisatellite probes

6.3.1 Advantages and disadvantages

Typing DNA using hybridization conditions under which hypervariable minisatellite loci are detected singly has the advantages over multilocus DNA fingerprinting that profiles are simpler to interpret, and require smaller amounts of DNA for high-quality results. The main disadvantage is that the amount of information per test is smaller—multiple locus-specific minisatellite probes need to be used consecutively to give resolving power approaching that of one DNA fingerprint. One compromise between the two, which gives higher information content per hour of work, but which complicates the interpretation somewhat, is to use pools of single-locus probes [3,5]. An important condition for obtaining results to which appropriate statistical weight can be attached is the use of loci for which allele frequencies (or at least mean or maximum allele frequencies) are known—the general properties of a number of widely used loci are shown in Table 6.3.

Protocol 11 describes the typing of individual

Locus	Probe	Enzyme	Allele frequency	Ref.
D1S7	pMS1	HinfI	0.04 (max.) 0.02	[36] [5]
D1S8	pMS32	AluI	0.03 (mean)	[5]
D2S44	YNH24	HinfI	0.05 (max.)	[36]
D7S21	pMS31	HinfI	0.02 (mean)	[37]
D7S22	p λ dag3	HinfI	0.03 (mean)	[5]
D14S13	CMM101	HinfI	0.11 (max.)	[36]
D17S79	pAC256	HaeIII	0.261 (max.)	[36]

Table 6.3 Properties of minisatellite loci.

minisatellite loci using locus-specific minisatellite probes.

6.3.2 Typing individual minisatellite loci: data interpretation and statistical evaluation

6.3.2.1 Match criteria

The simplest question in the interpretation of profiles from single minisatellite loci concerns band matching. In principle, the criterion is simple: bands of indistinguishable mobilities constitute a match, bands of discrepant mobilities constitute an exclusion. In practice, the problem is that even samples from an identical source, run side-by-side on a gel, may exhibit slightly different mobilities (Fig. 6.2). In DNA fingerprinting the issue is usually simple to resolve; profiles which are identical except for a small shift down the gel should be obvious.

This ‘band-shift’ effect may be due to a number of factors, including the ionic strength and degree of degradation of the sample; in Fig. 6.2b, the simplest explanation for the observed pattern is that the two samples are from different sources. It remains possible, however, that the two samples are from the same source, but that impurities or degradation in one sample has caused a co-ordinate shift in the positions of both bands.

In forensic practice, where DNA is often of poor quality and minuscule quantity, these effects can cause real problems of interpretation. In research, a number of simple criteria can be used to resolve the issue. Firstly, other loci used to type the filter can give corroborating evidence; either by giving an unambiguous exclusion, or by showing a similar pattern of a ‘shifted match’. Another simple expedient possible in research (but only rarely in forensic use) is to mix the samples and type the mixture.

6.3.2.2 Mutation or exclusion?

In paternity analysis, failure to match a non-maternal band with an alleged father may have one of two explanations: that the man is not father, or that a germline mutation has occurred on transmission from this man (the true father) to his offspring. For this reason the use of a single

minisatellite locus is not sufficient to exclude paternity. This is also the reason why the most informative loci, which can have germline mutation rates of 15% per sperm [27], are not useful in paternity analysis. The best loci are those which have large numbers of rare alleles (and hence very high heterozygosities) but have relatively low germline mutation rates (1% or lower). In practice, germline mutation can be simply distinguished from incorrect paternity by analysis of the same samples at further hypervariable loci.

6.3.2.3 Allele frequencies and probabilities

Once the issues surrounding matching bands have been resolved, the statistical weight of a given match can be ascertained. The first basic issue, when an allele (A) is matched, is whether to use the allele frequency of that specific allele (q_A) or to use a general ‘mean allele frequency’ from the locus as a whole. The advantage of the former approach is precision, but the drawback is that it requires both accurate sizing of the fragment (to identify the allele correctly) and a detailed allele frequency database (to assign the correct frequency to that allele). In these circumstances the known allele frequencies can be used to assign probabilities as discussed in Section 6.1.4.

The use of mean allele frequency to apply to all alleles at a locus is particularly appropriate for the most informative loci, at which there are large numbers of alleles (giving a quasi-continuous size distribution), all alleles are rare, and therefore drawing up an allele-by-allele frequency distribution is not realistic; the observed frequency distributions reflect the occupancy of size-classes of alleles [28]. At these loci [5,28] a good estimate of the mean allele frequency q can be obtained from the observed heterozygosity H by $q = (1 - H)$. Note that at these highly polymorphic loci the observed heterozygosity and mean allele frequency are dependent on the gel resolution conditions: the better the resolution of the gel, the better ‘close heterozygotes’ can be resolved, and thus the higher the observed heterozygosity and the lower the mean allele frequency. Using the observed *maximum* allele frequency (see Table 6.3) for q would lead to a conservative estimate of statistical significance. If using this method, then the probability of a chance match of both bands between unrelated samples is $2q^2$, and that of the chance presence of a non-maternal band in an alleged father is $2q - q^2$ (compare Section 6.1.4). Sequential use of five or six loci of high heterozygosity can give cumulative probabilities approaching the resolution given by DNA fingerprinting [5].

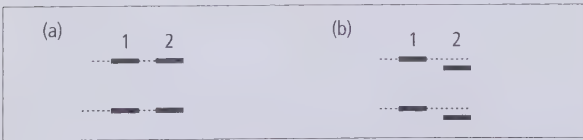


Fig. 6.2 DNA profiles. (a) Matching DNA profiles at a single polymorphic locus. (b) Discordant profiles, or just a ‘gel-shift’?

Locus-specific minisatellite probes can be used in pools to give a 'reduced DNA fingerprint'. Here, as with DNA fingerprinting itself, locus-specific information is lost (you don't know which band comes from which locus) and so the use of locus-specific allele frequencies is not appropriate. The empirically observed frequency of band-sharing between unrelated individuals (again, dependent upon gel resolution) can be used to derive probabilities for a given match. Pooled locus-specific probes are very powerful in resolving issues of identity, but are of limited resolution in paternity or twin analysis [5].

6.4 PCR analysis of minisatellites

6.4.1 AMPFLPs and MVR-PCR: advantages and disadvantages

This section will discuss typing methods which use PCR typing of minisatellites, either by direct amplification across minisatellite alleles to give amplified length polymorphisms (AMPFLPs, Fig. 6.3) or internal analysis of minisatellite variant repeats (MVRs) within an allele (Fig. 6.4). The main advantages of these methods lie in the combination of the high informativeness of minisatellites with the sensitivity of PCR, and for MVR-PCR the production of a digital diploid code which makes no assumptions about allele identity or frequency. The chief disadvantage of AMPFLPs is the importance of careful control of PCR conditions, to avoid spurious or incomplete profiles; for example, under inappropriate PCR conditions (in particular inadequate extension times), or if the DNA is badly degraded, the amplification of a long allele will be less

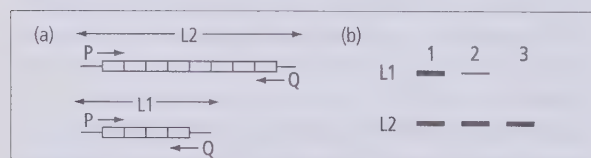


Fig. 6.3 PCR typing of a minisatellite locus. (a) Amplifying across the tandem repeat array between flanking primers P and Q should (in this heterozygous individual) give two fragments of lengths L1 and L2, corresponding to the two alleles. (b) If both alleles amplify efficiently under the conditions used, then the profile will be complete (lane 1); however, if the conditions discriminate against longer alleles (for example with inadequate extension times), then the longer product may be only faint (lane 2) or even invisible (lane 3). In this last case therefore there is a risk of mistyping this heterozygous individual as homozygous for the shorter allele.

favoured than that of a smaller allele, such that the longer allele may be faint or even disappear altogether—allele 'drop-out' (Fig. 6.3b). For MVR-PCR, the main disadvantage is that although extremely variable, only a single locus is being typed. This has the important consequences that resolution of sibs will be very poor (1:4), and that resolution of parentage will be very 'hit-or-miss' (see Section 6.4.3, below).

6.4.2 Available loci

The difficulties inherent in faithfully amplifying tandemly repeated, very GC-rich alleles of (for example) 5 kb, even given recent advances in the amplification of long templates [29,30], preclude the simple use of a typical hypervariable minisatellite locus for reliable DNA typing as AMPFLPs (for discussion, see [31]). For reliable typing, only the smaller minisatellite loci have so far proved useful. While this consideration rules out many of the most informative minisatellite loci, a number of highly informative systems have been developed; details of PCR typing have been published for the

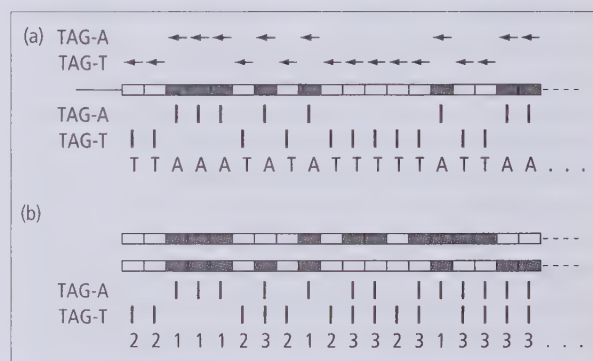


Fig. 6.4 General principles of MVR-PCR. (a) The interspersed pattern of variant repeat types ('A', shaded or 'T', white) at one end of a minisatellite allele can be mapped by doing two PCR reactions. Amplifying between the fixed flanking primer 'O' and TAG-A will produce a series of products, the lengths of which correspond to the positions of the 'A' type repeat units. Similarly, using TAG-T and primer O will produce a series of products whose lengths correspond to the positions of the 'T' type repeats. Running the products from each reaction side-by-side on a gel will then allow the sequence of repeat unit types at that end of the allele to be read. (b) Typing diploid genomic DNA using the same primers will give a pattern resulting from the superimposition of the patterns from each of the two alleles. Thus at each position one of three basic results is possible: both alleles have an 'A' type repeat (code 1), both alleles have a 'T' type repeat (code 2), or the alleles have different repeat types at this position (A + T = code 3). For further details see refs 43–46.

minisatellites at COL2A1 [32,33], ApoB [34], D17S5 [35], RB1 [36], D17S30 [37] and D1S80 [38]. A system for fluorescent typing of minisatellites in multiplex [39] has also been developed.

The choice for MVR-PCR is yet more restricted. Digital typing of diploid DNA by MVR-PCR requires that the minisatellite is composed of repeat units of the same length, such that the 'ladder' of products from each allele do not lose registration with one another. For this and other technical reasons, although a number of MVR typing systems have been developed for research into the mechanisms of minisatellite instability [40–42], D1S8 remains the only locus at which simple MVR-typing of diploid DNA can be performed [43–46].

Protocol 12 describes simple (two-state) MVR-PCR as modified from the original description [43]. Further informativeness can be obtained by the 'four-state' mapping described, with slight modification of the PCR conditions [45].

6.4.3 Data interpretation and statistical evaluation

6.4.3.1 AMPFLPs

Since both techniques deal with alleles at individual loci, the statistical interpretation of profiles produced by AMPFLP analysis is the same as for single minisatellite loci typed by Southern blot hybridization (see Section 6.3.2). Briefly, if alleles can be sized with sufficient precision to allow unambiguous identification, then the frequencies of individual alleles can be used directly in the calculation of probabilities; alternatively, the frequency of all alleles can be conservatively set at the maximum allele frequency recorded for any allele at the locus.

6.4.3.2 MVR-PCR

Unlike AMPFLPs, MVR-PCR gives a diploid phenotype in the form of a ternary code. The match criterion is simple: if the codes match, the samples match. The probability of this occurring purely by chance can be conservatively estimated at $3/N$ (with 95% confidence), where N is the total number of recorded codes to date, in which the sample code is not found; N presently stands at 625 unrelated individuals (A.J. Jeffreys *et al.* unpublished). It is important, however, to remember the limitations of MVR-PCR at a single locus: the probability above refers to comparisons between unrelated individuals—the corresponding probability between sibs is (approximately) 0.25. Furthermore, in assessing parentage its high informativeness is offset by the high frequency of new mutations; if the data are consistent with parentage as stated, then the data

can give strong support. However, if there are discrepancies between the MVR codes of a child and one of the parents, it is not possible to distinguish simply between germline mutation at D1S8 (frequency about 1%) and simple incorrect parentage [43].

6.5 PCR typing using simple tandem repeats

6.5.1 Advantages and disadvantages

Typing simple tandem repeat (STR) loci has the great advantage that many laboratories doing research in human genetics will be familiar with the methodology; indeed, as mentioned above (Section 6.1.2), many of the samples in question will have been genotyped as part of, for example, linkage studies, and thus data on these loci will already be available with no extra practical work. As with PCR methods in general, small amounts of DNA are required, and the small size of the products means that most alleles can be classified by size with precision. Although STRs are generally less informative than minisatellite loci, this is related to the relative germline stability of STR loci; complications caused by germline mutations are therefore very unlikely. The main disadvantage is that there is much less information per locus, since at even the most informative STR loci some alleles can have high population frequencies—for example, 0.2 or more. In practice, this disadvantage can be offset by the use of many different loci (see Section 6.5.3).

6.5.2 Typing STR loci

Methods for the typing of di-, tri- and tetranucleotide repeat loci have been discussed in Sections 5.3.5.1 and 5.3.5.2 of Chapter 5. In general, tri- or tetranucleotide repeat loci give 'cleaner' results than dinucleotides. Given the choice, it pays to use loci for which large data sets have been tested already, so that reasonably accurate estimates of the allele frequencies are available [47–49]. Having said that, many studies will involve large numbers of dinucleotide loci, and the estimates of allele frequencies deduced during the initial characterization of the locus can be used, although in many cases these will not have been based on very large surveys.

6.5.3 Data interpretation and statistical evaluation

Since simple repeat loci usually have discrete allelic states which can be identified unambiguously by

accurate sizing, match criteria (see Section 6.3.2) can be unambiguous. The frequencies of individual alleles can be obtained from the Genome Data Base (GDB, see Chapter 37 for address), or the original locus descriptions, or by further characterization of the locus. The calculations of statistical weight can then follow the simple principles outlined in Section 6.1.4.

It has been a matter of considerable debate in the legal uses of DNA typing whether information from

unlinked loci can really be treated as independent, or whether considerations of population genetics suggest that information from different loci may show some interdependence [14–19]. For pure research purposes, where nobody's liberty is at stake (Section 6.1.5), it is a safe approximation to treat information from different loci as independent, and thus simply to multiply probabilities from individual loci to give the overall probability of a set of results.

Protocol 10 DNA fingerprinting

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- genomic DNA
- restriction enzyme (*Hinfl* or *Alul*)
- 20–30 cm agarose gel for electrophoresis
- 0.5× TBE buffer: To make 10× TBE buffer (per litre): 109 g Tris base, 55 g boric acid, 9.3 g disodium EDTA
- nylon or nitrocellulose filters
- 10× Denhardt's solution: 0.2% bovine serum albumin /0.2% Ficoll 400/0.2% polyvinylpyrrolidone (MW approx. 44 000). This can be prepared as a 100× stock. Store at –20 °C.
- (alternative) phosphate/SDS hybridization solution: 0.5 M sodium phosphate (pH 7.2), 7% SDS, 1 mM EDTA, SSC
- SDS
- PEG 6000
- sheared, denatured herring sperm DNA (for the preparation of sheared and denatured DNA, see auxiliary protocol to Protocol 6 in Chapter 5).

Method

- 1** Digest 5–10 µg genomic DNA with a restriction enzyme which cuts very frequently, such as *Hinfl* or *Alul*.
- 2** Separate the fragments by electrophoresis on a long (20–30 cm) 0.8% agarose gel, until marker fragments of 2 kb are near the end of the gel. Our experience suggests that gels that are run slowly, in 0.5× TBE buffer, give the best resolution of fragments. Gels can be run, for example, over 48 h, with a change of buffer after the first 24 h.
- 3** Blot the DNA onto nylon (or nitrocellulose) filters. Fix the DNA to the membrane (with UV for nylon, baking for nitrocellulose).

- 4 Prehybridize sequentially in the following solutions at 65°C (at least 30min each change):
 - 10×Denhardt's solution/2×SSC/0.1%SDS;
 - 10×Denhardt's/2×SSC/0.1% SDS/6% polyethylene glycol 6000;
 - 10×Denhardt's/2×SSC/0.1% SDS/6% polyethylene glycol 6000.As a simpler alternative, prehybridize in phosphate/SDS hybridization solution at 65 °C (at least 30 min).
- 5 Probe preparation: after labelling the hybridization probe (see discussion below), purify the labelled fragment away from the unincorporated dNTPs.
- 6 Hybridize in 10×Denhardt's/2×SSC/0.1% SDS/6% polyethylene glycol 6000 (or, if using this method, phosphate/SDS solution), overnight at 65 °C.
- 7 Wash filters in 2×SSC/0.1% SDS at 65 °C.

The use of frequently cutting restriction enzymes is an important factor in getting clear multilocus profiles; these enzymes leave little flanking DNA with each tandemly repeated array, such that the size of the restriction fragment is determined almost entirely by the size of the repeat block. Most highly polymorphic fragments in the profile will be larger than 2 kb, and are thus best resolved by the extended electrophoresis suggested in step 2. Blotting onto nylon gives perfectly acceptable profiles; nitrocellulose is less convenient to handle, but can also be used.

There is great diversity in the possible methods for probe labelling; this will in part be dictated by the type of probe available. Probes prepared by standard random oligonucleotide primer labelling techniques [50] work fairly well, but additional specific activity can be obtained from specifically primed probes [1,51] or riboprobes [52]. Many of these procedures have also been adapted to incorporate non-radioisotopic labels. Very different probe preparation (and washing) conditions apply in the case of synthetic oligonucleotides such as (CAC)_n [58]. The posthybridization washing must, of course, be at low stringency.

If too many different loci are detected (see below), cross-hybridization can be generally suppressed by adding sheared denatured herring sperm DNA, which in these systems can act as a competitor rather than a blocking agent [59], to a final concentration of 50 µg ml⁻¹. However, if a particular probe/species combination has not been tried before, it is best to start by doing the hybridization (as in the protocol above) without competitor DNA. In new systems, wash carefully, beginning at very low stringency (say, 2×SSC, 0.1% SDS at 60 °C). Use a hand-held monitor to assess the distribution of retained radioactivity; relatively short minisatellite loci are much more numerous than longer ones, and thus good DNA fingerprints will have most detectable signal in the *smaller* fragments [2,20].

Troubleshooting

No signal/faint signal

- Check the incorporation of radioactivity into the probe. The method will depend on the type of probe used; for probes other than oligonucleotides, check a small amount of the recovered probe by scintillation counting. Estimate the specific activity of the probe. Probes with specific activities of 5×10^8 d.p.m. μg^{-1} or more should work well.
- Check the amounts and condition of the genomic DNA used.
- Try washing at lower stringency, for example $2 \times \text{SSC}$ at 60°C .

Too many bands

Occasionally so many different loci are detected that the profile is too ‘crowded’, with the bands forming an almost continuous smear. Fewer bands will allow greater resolution.

- Wash at higher stringency, for example, $1 \times \text{SSC}$ at 65°C .
- Add competitor (sheared and denatured herring DNA) to $50 \mu\text{g ml}^{-1}$ during the hybridization. For the preparation of sheared and denatured DNA, see auxiliary protocol to Protocol 6 (Chapter 5).

It is in the nature of the cross-hybridization underlying DNA fingerprinting that many different tandemly repeated sequences can, under the right conditions, detect multiple polymorphic loci. One spectacular example of this is the use of entirely synthetic tandem repeat probes to detect polymorphic loci in human DNA, some of which produce a DNA fingerprint, some of which give a single-locus profile, and some of which detect a few loci [27].

Another consequence of the same consideration is the ability of probes from one species to be useful in another; the probes are not (usually) detecting cognate loci [53], but fortuitously cross-hybridizing with different sets of polymorphic loci in the two genomes. Thus probes which detect single loci in human DNA may also have applications in non-human species. Some of these show informative multilocus (DNA fingerprinting) patterns [25,26,54], while others may sometimes fortuitously detect a single predominant locus [54]. This last result has a parallel in the behaviour of synthetic tandem arrays [27], and may be a labour-saving route to the development of single locus minisatellite probes, particularly from non-human species.

.....

Protocol 11 Typing using locus-specific minisatellite probes

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- genomic DNA
- restriction enzyme (*HaeIII*, *HinfI* or *AluI*)
- 20–30 cm agarose gel for electrophoresis
- nylon or nitrocellulose filters
- phosphate/SDS hybridization solution: 0.5 M sodium phosphate (pH 7.2)/7% SDS /1 mM EDTA
- SSC
- SDS

Method

- 1 Digest 1–10 µg genomic DNA with a restriction enzyme which cuts very frequently, such as *HaeIII*, *HinfI* or *AluI*. Use an enzyme known to be compatible with all the probes to be used.
- 2 Separate the fragments by electrophoresis on a long (20–30 cm) 0.8% agarose gel; run the gel as far as possible to achieve maximum resolution, but still retain the smallest fragments from any of the loci under study.
- 3 Blot the DNA onto nylon (or nitrocellulose) filters. Fix the DNA to the membrane (with UV for nylon, baking for nitrocellulose).
- 4 Pre-hybridize in phosphate/SDS solution at 65 °C.
- 5 Label the hybridization probe by random priming, and purify the labelled fragment away from the unincorporated dNTPs.
- 6 Hybridize overnight at 65 °C in phosphate/SDS solution.
- 7 Wash filters in 0.1 × SSC/0.01% SDS at 65 °C.

.....

Troubleshooting

Not enough signal

Most highly polymorphic minisatellite probes are ‘good’ probes, in the sense that they give good signals on autoradiography from a given amount of DNA. While it is possible to type submicrogram amounts of genomic DNA using these probes [5], you may have to wait for a long exposure before seeing the signal. If the amount of genomic DNA used is definitely not limiting, then check all the obvious things. Was the right restriction enzyme used? Was the probe DNA the size reported? How much probe DNA was labelled? Did it label well?

Too much signal/ complex profile

If filters are washed at too low a stringency, then an individual's DNA can give multiple hybridizing bands instead of the one or two traditional for a single locus. What is happening here is that by cross-hybridizing to other loci of similar sequence, a 'DNA fingerprint' has inadvertently been produced [5]. This might incidentally produce valuable additional information, but to reduce the complexity of the profile, wash at higher stringency: 0.1×SSC/0.01% SDS at 65°C usually leaves only signal from the cognate locus. Some of the cloned DNA probes used for typing single minisatellite loci may include some sequences from a nearby dispersed repeat, and may thus cause a non-specific 'smear' of hybridization in addition to the main bands. This can be suppressed by including denatured human DNA as a competitor in the hybridization [5].

Protocol 12 MVR-PCR at D1S8

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- genomic DNA
- PCR buffer [55]: 0.45 M Tris-HCl (pH 8.8), 110 mM (NH₄)₂SO₄, 45 mM MgCl₂, 67 mM β-mercaptoethanol, 45 μM EDTA, 10 mM each dNTP, 1.1 mg ml⁻¹ BSA (see note in 'Troubleshooting' for this protocol)
- Taq DNA polymerase
- MVR-PCR primers:
32-TAGA 5' TCATGCGTCCATGGTCCGGACATTCTGAGTCACCCCTGGC 3'
32-TAGT 5' TCATGCGTCCATGGTCCGGACATTCTGAGTCACCCCTGGT 3'
TAG 5' TCATGCGTCCATGGTCCGGA 3'
32-O 5' GAGTAGTTTGGTGGGAAGGGTGGT 3'
- 20–30 cm agarose gel for electrophoresis
- 0.5×TBE buffer: to make 10×TBE buffer (per litre): 109 g Tris base, 55 g boric acid, 9.3 g disodium EDTA
- nylon filters
- MS32 hybridization probe
- hybridization solution: 0.5 M sodium phosphate (pH 7.2)/7% SDS/1 mM EDTA
- SSC
- SDS

Method

1 For each sample to be typed, set up two PCRs as follows, one

containing 32-TAGA, the other 32-TAGT:

- 1 µl 20–100 ng DNA;
- 1 µl PCR buffer;
- 0.4 µl 10 mM primer 32-O;
- 0.4 µl 10 mM primer 32-TAG;
- 0.1 µl *Taq* DNA polymerase (5 U ml⁻¹);
- 5 µl H₂O;

and

- 1 µl 10 mM primer 32-TAGA;

or

- 1 µl 20 mM primer 32-TAGT;
(total, approx. 10 µl.)

Cycle at

96 °C for 50 s

69 °C for 45 s

70 °C for 3 min, 5 cycles, followed by

96 °C for 50 s

66 °C for 45 s

70 °C for 3 min, 17 cycles.

- 2 Run the products on a 1.2% agarose gel in 0.5×TBE buffer; put the two reactions ('A' and 'T') from one sample in adjacent lanes. The best resolution will be obtained using a long (30 cm) gel.
- 3 Blot onto a nylon filter, 2–4 h; fix the DNA to the filter by UV crosslinking.
- 4 Prehybridize filters in 20 ml phosphate/SDS buffer for 0.5–1 h at 65 °C.
- 5 Label 10 ng of MS32 probe by oligo-labelling [50].
- 6 Discard the prehybridization buffer, and replace it with 20 ml fresh phosphate/SDS buffer. Add the (boiled) MS32 probe and hybridize overnight at 65 °C.
- 7 Wash in 0.1 × SSC, 0.01% SDS, 65 °C, and autoradiograph.

.....

Troubleshooting

The profiles obtained should produce an even distribution of signal in bands extending for at least 50 repeats into the alleles. Examples of good quality results can be found in refs 43–46.

Not enough signal

The method is highly sensitive and should produce very strong signals after autoradiography, even from very small amounts (20 ng or more) of input DNA. One possibility if signals are weak is that impurities in the DNA are inhibiting the PCR, with the solution that better results will be

obtained using less input DNA. It is also a very curious (but well-attested) observation that the source of the BSA in the PCR buffer is important; acetylated BSA, for reasons which are entirely unclear, gives poor results [56]. As a simpler alternative to detection by blotting and hybridization as described above (although the number of positions resolved will be fewer), amplification for 30 cycles under modified conditions can give enough product to visualize directly on a gel after ethidium staining [57].

Most signal in smallest fragments

Poor results can be obtained not because of low total signal, but poor representation of the longer fragments.

- To remedy this, use a lower concentration of the TAGA and TAGT primers at step 1; try a two- to fivefold reduction.

Acknowledgements

I am very grateful to Professor Sir Alec Jeffreys for his kind help with this chapter and for permission to include protocols developed by him. My thanks are also due to Rita Neumann, Darren Monckton, Keiji Tamaki, Annette MacLeod, Maxine Allen and David Neil for their input to the MVR-PCR protocols, and to Esther Signer for her comments on the DNA fingerprinting section. I gratefully acknowledge the support of the Wellcome Trust in my research.

References

1 Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985) Hyper-variable 'minisatellite' regions in human DNA. *Nature* **314**, 67–73.

2 Jeffreys, A.J., Wilson, V. & Thein, S.L. (1985) Individual-specific 'fingerprints' of human DNA. *Nature* **316**, 76–79.

3 Jeffreys, A.J., Turner, M. & Debenham, P. (1991) The efficiency of multilocus DNA fingerprint probes for individualization and establishment of family relationships, determined from extensive casework. *Am. J. Hum. Genet.* **48**, 824–840.

4 Jeffreys, A.J., Brookfield, J.F.Y. & Semeonoff, R. (1985) Positive identification of an immigration test-case using human DNA fingerprints. *Nature* **317**, 818–819.

5 Wong, Z., Wilson, V., Patel, I. *et al.* (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.* **51**, 269–288.

6 Telenius, H., Clark, J., Marcus, E. *et al.* (1990) Minisatellite DNA profiles: rapid sample identification in linkage analysis. *Hum. Hered.* **40**, 77–80.

7 Hill, A.V.S. & Jeffreys, A.J. (1985) Use of minisatellite DNA probes for determination of twin zygosity at birth. *Lancet* **ii**, 1394–1395.

8 Helmuth, R., Fildes, N., Blake, E. *et al.* (1990) HLA-DQ α allele and genotype frequencies in various human populations, determined by using enzymatic amplification and oligonucleotide probes. *Am. J. Hum. Genet.* **47**, 515–523.

9 Gill, P., Ivanov, P., Kimpton, C. *et al.* (1994) Identification of the remains of the Romanov family by DNA analysis. *Nature Genet.* **6**, 130–135.

10 Tuncbilek, E. & Koc, I. (1994) Consanguineous marriage in Turkey and its impact on fertility and mortality. *Ann. Hum. Genet.* **58**, 321–329.

11 Lander, E.S. (1989) DNA fingerprinting on trial. *Nature* **339**, 501–505.

12 Koshland, D.E. Jr (1992) Editorial: DNA fingerprinting and eyewitness testimony. *Science* **256**, 593.

13 Thompson, W.C. & Ford, S. (1991) The meaning of a match: sources of ambiguity in the interpretation of DNA prints. In *Forensic DNA Technology* (Farley, M.A. & Harrington, J.J., eds) 93–152 (Lewis Publishers, Inc., Chelsea, Michigan).

14 Lander, E.S. (1991) Invited editorial: Research on DNA typing catching up with courtroom application. *Am. J. Hum. Genet.* **48**, 819–823.

15 Cohen, J.E. (1990) DNA fingerprinting for forensic identification: potential effects on data interpretation of subpopulation heterogeneity and band number variability. *Am. J. Hum. Genet.* **46**, 358–368.

16 Lewontin, R.C. & Hartl, D.L. (1991) Population genetics in forensic DNA typing. *Science* **254**, 1745–1750.

17 Chakraborty, R. & Kidd, K.K. (1991) The utility of DNA typing in forensic work. *Science* **254**, 1735–1739.

18 Brookfield, J. (1992) Law and probabilities. *Nature* **355**, 207–208.

- 19 Lander, E.S. & Budowle, B. (1994) DNA fingerprinting dispute laid to rest. *Nature* **371**, 735–738.
- 20 Jeffreys, A.J., Wilson, V., Thein, S.L. *et al.* (1986) DNA 'fingerprints' and segregation analysis of multiple markers in human pedigrees. *Am. J. Hum. Genet.* **39**, 11–24.
- 21 Royle, N.J., Clarkson, R.E., Wong, Z. & Jeffreys, A.J. (1988) Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics* **3**, 352–360.
- 22 Armour, J.A.L., Wong, Z., Wilson, V. *et al.* (1989) Sequences flanking the repeat arrays of human minisatellites: association with tandem and dispersed repeat elements. *Nucleic Acids Res.* **17**, 4925–4935.
- 23 Wong, Z., Wilson, V., Jeffreys, A.J. & Thein, S.L. (1986) Cloning a selected fragment from a human DNA 'fingerprint': isolation of an extremely polymorphic minisatellite. *Nucleic Acids Res.* **14**, 4605–4616.
- 24 Wells, R.A., Green, P. & Reeders, S.T. (1989) Simultaneous genetic mapping of multiple human minisatellite sequences using DNA fingerprinting. *Genomics* **5**, 761–772.
- 25 Georges, M., Lathrop, M., Hilbert, P. *et al.* (1990) On the use of DNA fingerprints for linkage studies in cattle. *Genomics* **6**, 461–474.
- 26 Julier, C., De Gouyon, B., Georges, M. *et al.* (1990) Minisatellite linkage maps in the mouse by cross-hybridization with human probes containing tandem repeats. *Proc. Natl Acad. Sci. USA* **87**, 4585–4589.
- 27 Vergnaud, G., Mariat, D., Apiou, F. *et al.* (1991) The use of synthetic tandem repeats to isolate new VNTR loci: cloning of a human hypermutable sequence. *Genomics* **11**, 135–144.
- 28 Budowle, B., Giusti, A.M., Waye, J.S. *et al.* (1991) Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. *Am. J. Hum. Genet.* **48**, 841–855.
- 29 Barnes, W.M. (1994) PCR amplification of up to 35kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proc. Natl Acad. Sci. USA* **91**, 2216–2220.
- 30 Cheng, S., Fockler, C., Barnes, W. & Higuchi, R. (1994) Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl Acad. Sci. USA* **91**, 5695–5699.
- 31 Jeffreys, A.J., Wilson, V., Neumann, R. & Keyte, J. (1988) Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Res.* **16**, 10953–10971.
- 32 Berg, E.S. & Olaisen, B. (1993) Characterization of the COL2A1 VNTR polymorphism. *Genomics* **16**, 350–354.
- 33 Wilkin, D.J., Koprivnikar, K.E. & Cohn, D.H. (1993) Heteroduplex analysis can increase the informativeness of PCR-amplified VNTR markers: application using a marker tightly linked to the COL2A1 gene. *Genomics* **15**, 372–375.
- 34 Boerwinkle, E., Xiong, W., Fourest, E. & Chan, L. (1989) Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction: application to the apolipoprotein B 3' hypervariable region. *Proc. Natl Acad. Sci. USA* **86**, 212–216.
- 35 Horn, G.T., Richards, B. & Klinger, K.W. (1989) Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nucleic Acids Res.* **17**, 2140.
- 36 Scharf, S.J., Bowcock, A.M., McClure, G. *et al.* (1992) Amplification and characterization of the retinoblastoma gene VNTR by PCR. *Am. J. Hum. Gen.* **50**, 371–381.
- 37 Rand, S., Puers, C., Skowasch, K. *et al.* (1992) Population genetics and forensic efficiency of 4 AMPFLPs. *Int. J. Leg. Med.* **104**, 329–333.
- 38 Budowle, B., Chakraborty, R., Giusti, A.M. *et al.* (1991) Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *Am. J. Hum. Genet.* **48**, 137–144.
- 39 Tully, G., Sullivan, K.M. & Gill, P. (1993) Analysis of six VNTR loci by multiplex PCR and automated fluorescent detection. *Hum. Genet.* **92** (6), 554–562.
- 40 Armour, J.A.L., Harris, P.C. & Jeffreys, A.J. (1993) Allelic diversity at minisatellite MS205 (D16S309): evidence for polarized variability. *Hum. Mol. Genet.* **2**, 1137–1145.
- 41 Neil, D.L. & Jeffreys, A.J. (1993) Digital DNA typing at a second hypervariable locus by minisatellite variant repeat mapping. *Hum. Mol. Genet.* **2**, 1129–1135.
- 42 Buard, J. & Vergnaud, G. (1994) Complex recombination events at the hypermutable minisatellite CEB1 (D2S90) *EMBO J.* **13**, 3203–3210.
- 43 Jeffreys, A.J., MacLeod, A., Tamaki, K. *et al.* (1991) Minisatellite repeat coding as a digital approach to DNA typing. *Nature* **354**, 204–209.
- 44 Monckton, D.G., Tamaki, K., MacLeod, A. *et al.* (1993) Allele-specific MVR-PCR analysis at minisatellite D1S8. *Hum. Mol. Genet.* **2**, 513–519.
- 45 Tamaki, K., Monckton, D.G., MacLeod, A. *et al.* (1993) Four-state MVR-PCR: increased discrimination of digital DNA typing by simultaneous analysis of two polymorphic sites within minisatellite variant repeats at D1S8. *Hum. Mol. Genet.* **2**, 1629–1632.
- 46 Tamaki, K., Monckton, D.G., MacLeod, A. *et al.* (1992) Minisatellite variant repeat (MVR) mapping: analysis of 'null' repeat units at D1S8. *Hum. Mol. Genet.* **1**, 401–406.
- 47 Urquhart, A., Oldroyd, N.J., Kimpton, C.P. & Gill, P. (1995) Highly discriminating heptaplex short tandem repeat PCR system for forensic identification. *BioTechniques* **18**, 116.
- 48 Hammond, H.A., Jin, L., Zhong, Y. *et al.* (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am. J. Hum. Genet.* **55**, 175–189.
- 49 Alford, R.L., Hammond, H.A., Coto, I. & Caskey, C.T. (1994) Rapid and efficient resolution of parentage by amplification of short tandem repeats. *Am. J. Hum. Genet.* **55**, 190–195.
- 50 Feinberg, A.P. & Vogelstein, B. (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**, 266–267.
- 51 Ferrie, R.M., Smith, H., Downes, E.A. *et al.* (1991) Repeat unit multipriming and hybridization—a novel method for the production of DNA fingerprints using minisatellite probes. *Nucleic Acids Res.* **19**, 2505.

- 52 Carter, R.E., Wetton, J.H. & Parkin, D.T. (1989) Improved genetic fingerprinting using RNA probes. *Nucleic Acids Res.* **17**, 5867.
- 53 Gray, I.C. & Jeffreys, A.J. (1991) Evolutionary transience of hypervariable minisatellites in man and the primates. *Proc. R. Soc. Lond. B.* **243**, 241–253.
- 54 Signer, E.N. & Jeffreys, A.J. (1993) Application of human minisatellite probes to the development of informative DNA fingerprints and the isolation of locus-specific markers in animals. In *DNA Fingerprinting: State of the Science* (Pena, S.D.J., Chakvabarty, R., Epplen, J.T. & Jeffreys, A.J. eds) (Birkhauser, Basel).
- 55 Jeffreys, A.J., Neumann, R. & Wilson, V. (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell* **60**, 473–485.
- 56 McKeown, B.J. (1994) An acetylated (nuclease-free) bovine serum albumin in a PCR buffer inhibits amplification. *BioTechniques* **17**, 246–248.
- 57 Yamamoto, T., Tamaki, K., Kojima, T. *et al.* (1994) DNA typing of the D1S8 (MS32) locus by rapid detection minisatellite variant repeat (MVR) mapping using polymerase chain reaction (PCR) assay. *Forensic Sci. Int.* **66**, 69–75.
- 58 Ali, S., Muller, C.R. & Epplen, J.T. (1986) DNA fingerprinting by oligonucleotide probes specific for simple repeats. *Hum. Genet.* **74**, 239–243.
- 59 Vassart, G., Georges, M., Monsieur, R. *et al.* (1987) A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. *Science* **235**, 683–684.
- 60 Vergnaud, G. (1989) Polymers of random short oligonucleotides detect polymorphic loci in the human genome. *Nucleic Acids Res.* **17**, 7623–7630.
- 61 Fowler, S.J., Gill, P., Werrett, D.J. & Higgs, D.R. (1988) Individual specific DNA fingerprints from a hypervariable region probe: alpha globin 3'HVR. *Hum. Genet.* **79**, 142–146.
- 62 Jeffreys, A.J., Royle, N.J., Wilson, V. & Wong, Z. (1988) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* **332**, 278–281.

Section 2

Physical mapping 1: cytogenetic analysis

Section 2 Introduction

Bryan D. Young

ICRF Department of Molecular Oncology, St Bartholomew's Hospital, London EC1A 7BE, UK

Cytogenetic analysis has been crucial in establishing the underlying genetic basis of many human diseases. The identification of the chromosomal defect associated with a disease has in many cases led to the identification of the affected genes. In certain genetic diseases, where the lesion is normally invisible cytogenetically, a rare form of translocation can occur, pinpointing the defective gene. The identification of the genetic basis of such diseases has led to new forms of diagnosis and may lead ultimately to novel therapeutic strategies. Cytogenetics has been particularly important in cancer research where many reciprocal translocations occur and have been used to identify the oncogenes involved. This in turn has led to valuable diagnostic tests based on polymerase chain reaction (PCR) analysis of gene fusions. In this way, minimal residual disease can be detected in patients undergoing therapy, thus complementing more traditional karyotype analysis.

Recent advances have expanded the utility and power of cytogenetic methods, in particular the development of fluorescence *in situ* hybridization (FISH). However, conventional chromosome analysis, as described in Chapter 7 (B. Czepulkowski), remains important. Most routine analysis still relies on Giemsa-banded karyotype analysis and in the hands of a skilled cytogeneticist subtle changes can often be detected. Alternative staining techniques

(quinacrine-, reverse-, and C-banding) also remain valuable in certain situations. The variations in methodology required for different sources of cells such as bone marrow, chorionic villus samples, and amniotic fluid are described.

The development of new fluorochromes has broadened the usefulness of FISH, as described in Chapter 9 (G. Senger and D. Sheer). In addition, a wide range of individual probes, ranging from yeast artificial chromosomes (YACs) to short cDNA fragments can now be used in hybridization studies using interphase nuclei, chromosomes or chromatin as targets. Digital microscopy, which has played an important role in the recording and interpretation of results (Chapter 13, N. Carter), is continuing to develop both in terms of hardware and software. FISH can be used not only for gene mapping but also for detection of abnormalities in clinical samples.

The use of whole chromosome probes or 'paints' has become increasingly popular (Chapter 10, L. Kearney) for analysing complex events beyond the scope of conventional banding. This technique usually relies on PCR labelling of a source of chromosome-specific DNA. Such starting material can be acquired either directly by flow sorting the chromosome of interest (Chapter 12, S. Monard) or by using a chromosome-specific library which is often derived from flow-sorted chromosomes. Flow

sorting has the advantage that highly purified single human chromosomes can be prepared, but has the disadvantage that chromosomes 9–12 cannot readily be separated from a normal human cell. This problem can be obviated, however, using monochromosomal cell hybrids. Flow sorting has also been used to prepare chromosomes that formed the basis of chromosome-specific DNA libraries. Such sources of enriched chromosomal DNA have been important in accelerating genome analysis.

Microdissection (Chapter 11, D. Lillington & A.N. Shelling) is another valuable technique for obtaining pure fragments of chromosomes. Although this method carries an inherent risk of contamination, it has the advantage that accurately dissected subchromosomal fragments can be obtained in a form

suitable for PCR amplification and cloning.

The application of all the above approaches has been particularly important in the study of the cancer cell. The greatest progress has been made in leukaemias and lymphomas where genetic changes tend to be simple and the tissue is readily accessible (Chapter 7). However, these methods are increasingly being applied to solid tumours (Chapter 8, S. Birdsall, Y.-J. Lu & J. Shipley), and a series of reciprocal translocations has now been used to identify the oncogenes involved. We can anticipate that further advances in dye technology and computer analysis will allow the development of even more advanced cytogenetic analysis. It is thus clear that cytogenetic analysis is an important and developing theme in human genetic analysis.



Chapter 7 Conventional cytogenetics

Barbara Czepulkowski

*Department of Cytogenetics/Haematology, Kings College Hospital, Denmark Hill, London
SE5 9RS, UK*



- 7.1 Introduction, 151**
- 7.2 Cell culture methodology, 152**
 - 7.2.1 Preparation of lymphocytes for chromosome analysis, 152**
 - 7.2.2 Prometaphase chromosomes, 152**
 - 7.2.3 Fragile X detection, 152**
 - 7.2.4 Prenatal diagnosis, 153**
 - 7.2.4.1 Amniotic fluid cultures, 153**
 - 7.2.4.2 Chorionic villus samples, 153**
 - 7.2.4.3 Culture maintenance for amniotic fluid and villus cultures, 153**
 - 7.2.4.4 Harvesting cultures, 153**
- 7.3 Banding techniques, 154**
- 7.4 Detection of constitutional abnormalities, 156**
 - 7.4.1 Approaches to analysis, 159**
- 7.5 Cancer cytogenetics, 159**
 - 7.5.1 Practical considerations, 161**
 - 7.5.2 Approaches to analysis, 167**
- 7.6 Digital imaging, 167**

Protocol 13
Processing of whole blood samples and slide preparation, 168

Troubleshooting, 170
Slides too crowded, interphases overpowering metaphases, 170

Slides very sparse, paucity of interphases and metaphases, 170
Metaphases broken up, otherwise known as ‘chromosome soup’, 170
Metaphases too clumped, cytoplasm still visible around metaphase plate, 170
Insufficient mitoses for analysis, 171

Protocol 14
Thymidine block synchronization method for obtaining prometaphase chromosomes, 171

Troubleshooting, 172
Insufficient or no metaphase spreads, 172

Protocol 15
Fragile X detection, 172

Protocol 16
Setting up amniotic fluid cultures, 173

Troubleshooting, 174
Infection occurs in culture vessels, 174

Protocol 17
Setting up a chorionic villus sample, 174

Protocol 18
Setting-up a long-term villus culture, 175

Troubleshooting, 175
Poor growth, 175
Maternal cell contamination, 176

Protocol 19

Direct villus culture, 176

Protocol 20

Harvesting amniotic fluid and chorionic villus cultures, 177

Troubleshooting, 178

Loss of material during processing, 178

Cells not floating off quickly during trypsinization, 178

Mitotic index poor, 179

Protocol 21

Giemsa banding, 179

Troubleshooting, 180

Chromosomes darkly stained, no real visible bands, 180

Chromosomes very fuzzy and puffed-up, 180

Chromosomes puffy, pale, ghost-like, bands indistinct, 180

Protocol 22

Quinacrine banding, 180

Troubleshooting, 181

Weakly fluorescing chromosomes, 181

High background fluorescence, 182

Protocol 23

Reverse banding, 182

Protocol 24

Constitutive heterochromatin banding, 183

Troubleshooting, 184

Barium hydroxide deposits on slide, 184

Protocol 25

Synchronization technique for bone marrow samples, 184

Acknowledgements, 185

References, 185

7.1 Introduction

Human cytogenetic analysis has been of great importance in characterizing genetic diseases, from the early discovery of trisomy 21 in Down’s syndrome to more recent localization of genes involved in disease. It has aided in the understanding of the origin and inheritance of such diseases, with the patients and their families benefiting from genetic counselling.

In addition, the past 20 years have seen a tremendous increase in the study of acquired chromosomal changes in malignancies (see Case Study 7.1 and Section 7.5). This chapter will address methods of obtaining chromosome preparations from blood, amniotic fluid and bone marrows, and the various banding techniques. The chapter is intended as a guide to the conventional techniques that give consistently useful results in a wide range of cytogenetic applications, and that are most commonly used for clinical diagnosis. Additional

information can be found in more specialized cytogenetics texts [1,2]. The chapter also deals with the main clinical applications of cytogenetics.

Although numerical chromosomal changes such

Cytogenetics is used for:

- evaluating karyotypes in mammalian and other species
- detection of prenatal abnormalities
- detection of acquired chromosome abnormalities in malignant disorders
- monitoring minimal residual disease in patients with malignant disorders
- assessing disease status in patients with malignant disorders
- detection of constitutional abnormalities in patients
- initial localization of disease genes
- mutagenicity testing and breakage syndromes
- detection of possible chromosome abnormalities in recurrent aborters

Applications box 7.1

The discovery of the *BCR-ABL* fusion gene resulting from the t(9;22) translocation

In cancer cytogenetics, research has been carried out and is still ongoing into the various translocations involved in disease. Cytogeneticists were the first to note that there were non-random chromosomal changes in a variety of tumours. More recently, involvement of proto-oncogenes has been indicated by observing that the loci of known oncogenes are close to certain breakpoints involved in the translocations. New oncogenes have been discovered by cloning and mapping certain malignancy-associated breakpoints in the chromosomes. In some cases, the translocations occur outside the protein-coding domain of the oncogene – for example, in the group of translocations associated with Burkitt’s lymphoma and B-cell acute lymphocytic leukaemia (ALL): t(8;14)(q24;q32), t(2;8)(p12;q24) and t(8;22)(q24;q11). In other cases, the rearrangement gives rise to hybrid genes, for example t(9;22) gives rise to the *BCR-ABL* fusion gene, whose discovery is described here.

The t(9;22) translocation, found in association with chronic phase myeloblastic leukaemia (CML), is an ideal example of how a hybrid oncogene may play an important role in causing a malignant disorder. In 1960, Nowell and Hungerford discovered a very consistent small chromosome in the cells of patients with CML, and called it the Philadelphia chromosome in honour of the city in which it was discovered [6]. Cytogenetically, the discovery of the Philadelphia chromosome sparked a great arousal of interest in chromosome changes in malignant disease. However, until it was possible to band chromosomes, cytogenetic changes appeared to be random in other diseases. When banding techniques were introduced in the 1970s, it became clear that the small Philadelphia chromosome was not in fact a deletion, but a product of

a translocation between chromosomes 9 and 22. The introduction of banding techniques also showed that cytogenetic changes were indeed non-random across a wide range of disorders.

The possibility of an oncogene being involved in the t(9;22) translocation first arose with the finding that the *ABL* oncogene, translocated from chromosome 9, mapped to an area close to the breakpoint on chromosome 22. Heisterkamp and colleagues were the first to clone the breakpoint region [7]. When a fragment representative of this region was cloned for further analysis, Groffen and colleagues proved they had isolated the exact translocation sites on 9 and 22 [8]. It was observed that the critical 22 breakpoint was localized to a small area, indicating that this region plays a vital role in CML. This suggested that the expression of *ABL* (or a closely linked gene) might be altered in some way during the translocation. At the same time, several groups reported the presence of an abnormal *ABL* mRNA in CML cells. This turned out to be a fusion mRNA derived partly from the gene on chromosome 22 called *BCR*, which stands for breakpoint cluster region, and partly from the translocated *ABL*.

Interestingly, an aberrant *BCR-ABL* fusion gene was observed in patients with ALL who have the t(9;22) translocation. Subsequent work showed that in ALL, a smaller fusion protein is consistently found, which is the result of a shorter *BCR* contribution to the 5’ end of the fusion mRNA. However, the two translocations look identical when observed cytogenetically, and are given the same breakpoints for both diseases. Intriguingly, following appropriate therapy, the translocation is eliminated in ALL patients, whereas in CML patients it tends to persist, although some success in eradicating the abnormal clone has been achieved with interferon therapy for CML patients.

Table 7.5 shows some of the other genes involved in various translocations.

as trisomies and gross rearrangements such as translocations are the usual abnormalities associated with cytogenetic study, the development of methods for preparing 'long' (prometaphase) chromosomes now enables the cytogeneticist to detect more subtle abnormalities, in particular small deletions and some translocations. Using banding methods other than the conventional Giemsa banding (G-banding) used by most laboratories, deletions can now be detected in pale G-banded regions (e.g. using reverse, or R-banding). Apart from protocols for banding techniques, this chapter covers the culture conditions required for different types of sample. In particular, special culture conditions are required if one is looking for a possible chromosomal fragile site or chromosomal instability syndrome.

7.2 Cell culture methodology

Blood cell karyotyping is the anchor of modern cytogenetics, as blood is the most accessible tissue and the growth potential of white blood cells following mitogen stimulation is normally very good. Using defined media, cultures can be established with 0.5ml of whole blood; slides are available for banding after 48–72 h of culture, and subsequent processing. Red cells and platelets in normal adult blood are not normally nucleated; only in early fetal life and in malignancy may nucleated red cells be present. Consequently, chromosome analysis is performed on the nucleated white cells (essentially the lymphocytes). For a more detailed explanation of the theory and history of cytogenetic analysis, the reader is referred to ref. 1. Phytohaemagglutinin (PHA) and Epstein–Barr virus (EBV) are the standard lymphocyte mitogens used in blood culture as they tend to give the most satisfactory results. EBV can also be used to transform cells to establish cell lines for research purposes (see Appendix II).

Blood samples are normally collected in sterile lithium heparin tubes and mixed gently to prevent clotting. A clotted sample is unacceptable for cytogenetic culture, although attempts can be made to disaggregate the clot by adding some preservative-free heparin, and rubbing the clot gently between sterile orange sticks. If blood arrives in an incorrect sample tube, such as one containing EDTA, the cells should be washed two or three times in medium (without serum) and then placed in the correct container before setting up in culture. A procedure for processing whole blood is given in Protocol 13. A 10-ml blood culture normally requires the following quantities of whole blood, depending upon the age

of the patient:	
adults and children over 5 years old	0.8 ml
children less than 5 years old	0.5 ml
infants up to 5 years old	0.1 ml
cord blood	0.3 ml
fetal blood	0.2 ml

7.2.1 Preparation of lymphocytes for chromosome analysis

Protocol 13 describes the processing of whole blood samples. Further information on handling and processing of blood samples is given in Appendix II.

7.2.2 Prometaphase chromosomes

Several methods are available for the production of 'long' prometaphase chromosomes that can be used for high-resolution banding. Protocol 14 gives good quality preparations, and is adapted from a method used in the Newcastle Northern Region Genetics Service (Department of Human Genetics, 19 Claremont Place, University of Newcastle upon Tyne, Newcastle NE2 4AA) which consistently produces excellent results. (A similar method is given in Protocol 58, Chapter 11 for synchronization of cultures with thymidine for the preparation of chromosomes for microdissection to make probes for microFISH.)

7.2.3 Fragile X detection

Recommendations from the International Fragile X Group [3] are that at least two different culture media should be used for neonatal and adult blood samples when screening for fragile X. This is because of the danger of false-negative results. The fragile X chromosome phenotype is most clearly expressed in conditions of folate deficiency and so the usual choices are a low-folate medium such as TC199 with 2% serum, FX-1 or Iscoves medium, and another with excess thymidine, and/or a medium with a folate antagonist such as metho-trexate (MTX). Medium with a low concentration of serum is required, as serum contains a small amount of folate and this would compromise the stringency requirements for expression of fragile X. In addition, a minimum of 100 cells is required for scoring. It is advisable to do family studies in cases of fragile X in order to assess familial sensitivity to folate deficiency. Protocol 15 describes a method for fragile X detection.

In any chromosome preparation, the cytogeneticist may encounter small numbers of chromosomes showing either gaps in chromatids or breaks

in the chromosome leaving a nonstaining gap. Most of these can occur randomly, but others occur at nonrandom sites which have been documented throughout the karyotype. These are normally considered to be inherited features of no known clinical significance. Indeed, the only clinically significant fragile site is the fragile X.

7.2.4 Prenatal diagnosis

The most commonly used tissue for prenatal diagnosis is amniotic fluid, although some laboratories use chorionic villus samples (CVS). Fetal blood may also be analysed using the blood culture methods described above (Protocol 13).

Removal of an amniotic fluid sample is known as *amniocentesis*, and is normally performed at about the 16th week of gestation. As cells have to be cultured long-term—that is, 2–3 weeks for amniocentesis samples and around 10 days for CVS cultures—aseptic tissue culture techniques must be adopted. Details of aseptic culture will be found in ref. 1 (see also Chapter 11, Protocol 59). The maintenance schedule for the cultures is given in Section 7.2.4.3 below and common procedures will also be discussed.

7.2.4.1 Amniotic fluid cultures

The amniotic fluid sample taken is usually 20 ml, preferably divided into two universal containers. In practice, less than this will often arrive in the laboratory, but can still be processed. It is important to keep a record of the condition of the sample on arrival—that is, volume, appearance (in particular whether blood is present in the sample) etc., and subsequent maintenance procedures, such as dates of medium change and harvest times. Protocol 16 describes the setting-up of amniotic fluid cultures.

7.2.4.2 Chorionic villus samples

Chorionic villus samples have been taken since the early 1980s for first trimester prenatal diagnosis. Spontaneously dividing cells are present in the cytotrophoblast layer of the villi, and are exploited in direct chorionic villus culture (see Protocol 19). However, it is always advisable also to perform a long-term culture, as the direct cultures are notoriously unreliable in providing sufficient metaphases of adequate quality for analysis. In long-term cultures, the mesenchyme core cells are analysed. Ten milligrams of villi from the chorion frondosum is usually adequate for both a long-term and direct culture. Table 7.1 shows the transport medium essential for the transport of chorionic villi from the place of sampling to the laboratory. This heparinized

Table 7.1 Chorionic villus sample transport medium.

Medium components	Volume (ml)
Basal medium, i.e. Ham's F10	100
FCS	10
L-Glutamine (200 mM)	1
Penicillin or streptomycin (10 000 IU ml ⁻¹ or 10 000 µg ml ⁻¹)	3
Kanamycin (10 000 µg ml ⁻¹)	3
Mycostatin (1000 IU ml ⁻¹)	0.3
Heparin (1000 IU ml ⁻¹)	1

medium prevents clotting when the aspirate is contaminated with maternal blood. In addition, the antibiotics prevent contamination by vaginal flora when the transcervical route is used for sampling.

Although villi can survive up to 3 days in the above transport medium, the chances of obtaining a direct result from samples experiencing such a delay are minimal. Such samples would normally be cultured long-term. Protocol 17 describes the setting up of a chorionic villus sample and Protocol 18 the procedure for long-term culture. Protocol 19 describes a method for direct villus culture.

7.2.4.3 Culture maintenance for amniotic fluid and villus cultures

Following undisturbed incubation for at least 6 days (5 in the case of chorionic villi), the cells should be examined using an inverted microscope, to assess cell growth. This can be variable, with some cultures growing quickly and others showing no signs of growth at this time. The medium should either be half or fully changed. Change the medium every two to three days, until harvesting is anticipated. If 10–14 days have elapsed and no growth is visible, the clinician should be alerted so that a repeat amniocentesis or chorionic villus sampling can be offered to the patient. In addition, Chang medium may be tried in cases where the culture was originally set up in Ham's F10. This may rescue a slow-growing culture.

7.2.4.4 Harvesting cultures

Protocol 20 describes harvesting of amniotic and chorionic villus cultures. For cultures growing in petri dishes, trypsinization is required to remove the cells from the substrate before harvesting. The advantage of *in situ* cultures on coverslips is that cells can be harvested without trypsinization. The cultures should be examined the day after a medium change for the presence of enough dividing cells (which have a rounded-up appearance). To these cultures 0.1 ml of colcemid solution (10 µg ml⁻¹) is

added for 2–4 h and incubated at 37°C. Following incubation, the culture tube should be checked again for large quantities of rounded, dividing cells.

7.3 Banding techniques

The method most commonly used in diagnostic laboratories is Giemsa banding (or G-banding) (Fig.7.1) using trypsin, which gives consistent

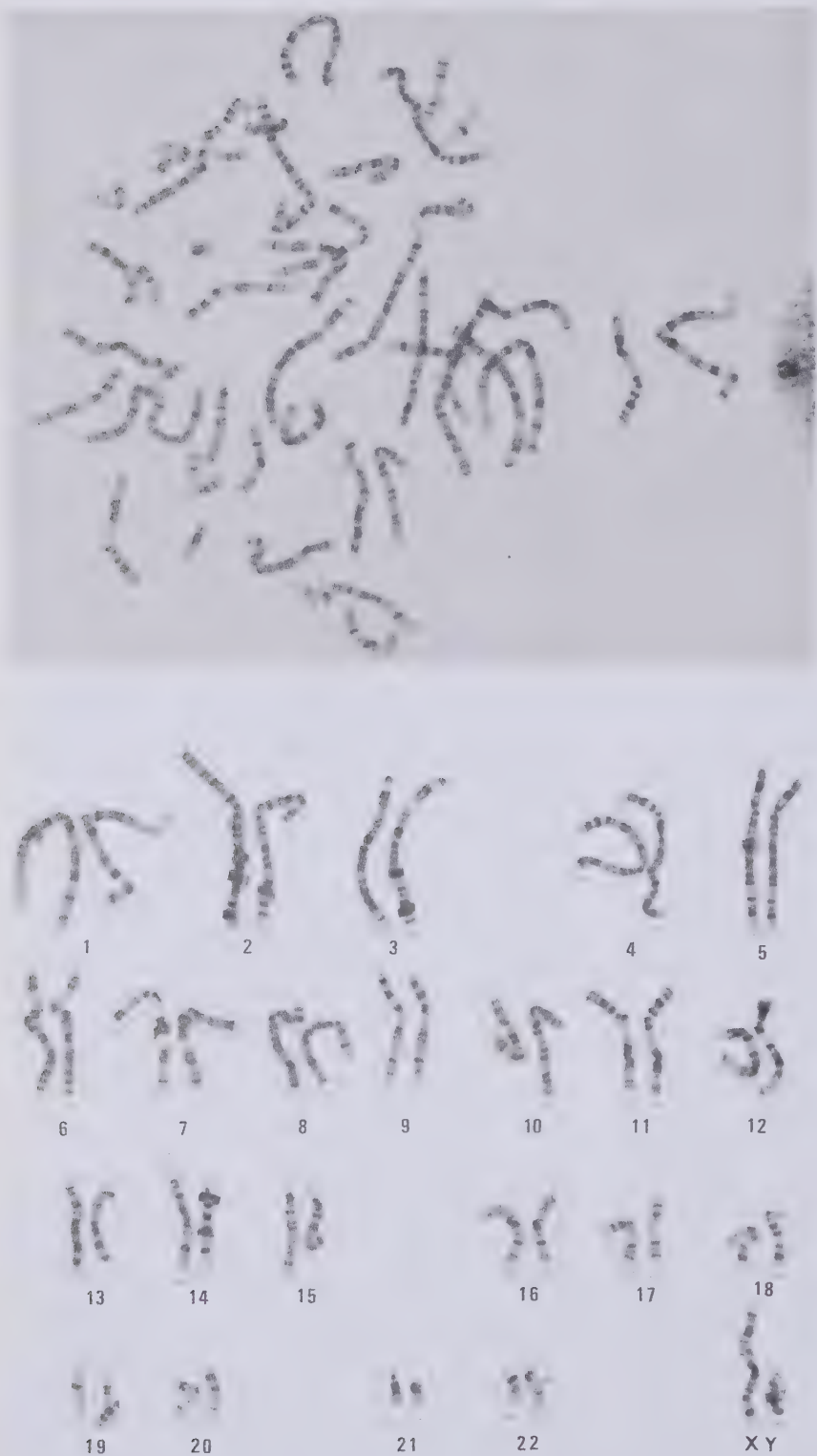


Fig. 7.1 G-banded cell (top) with karyotype (below) from a normal male, 46,XY. See Protocol 21. From [13] by permission of Oxford University Press and Dr P.A. Benn.

results. Trypsin treatment produces the characteristic banding pattern on the chromosomes, which is then visualized by staining with Giemsa, but the mechanism of banding is unknown or unclear. Although still called G-banding, similar results are also obtained with Leishman's stain, which is more suitable for bone marrow samples. Optimum results are obtained when slides are aged for 3–5 days, but in urgent cases, they can be left on a hotplate or in an oven at 56–60 °C overnight. For slides over 1 week old, a longer time in trypsin may be required. There

are numerous methods for obtaining G-bands. Protocol 21 gives consistent results.

Many more banding techniques exist, which are most often used for research purposes, and of course they vary with personal laboratory choice. For example, methods for G-banding using Wright's stain, and for achieving high-resolution early and late replication banding (equivalent to R- and G-banding, respectively) by BrdU incorporation are described in Chapter 9. Table 7.2 describes the more commonly used methods and their applications (see, e.g.

Table 7.2 Banding techniques and their applications.

Techniques	Applications
Giemsa banding (G-banding)	General chromosome recognition
Quinacrine banding (Q-banding)	As for G-banding but using a fluorescent stain
Constitutive heterochromatin banding (C-banding)	Staining of centromeres and heterochromatin. Used for examining polymorphisms between homologues and individuals, for familial markers, and marker chromosomes
Reverse banding (R-banding)	Staining of pale G-band areas
Nucleolar organizer region (NOR) staining	Detection of NOR regions, which contain 18S and 28S rRNA genes, on chromosomes 13, 14, 15, 21 and 22. It is used in delineating breakpoints in Robertsonian and reciprocal translocations (see Fig. 7.2)
Early and late differential replication banding (Chapter 9, Protocol 38)	Detection of different cycles of replication. Used to investigate how different parts of chromosomes replicate at different times in the cell cycle
DA/DAPI staining	Distamycin A (DA) and DAPI both have an affinity for AT base pairs, and bind at similar but not identical sites. Highlights heterochromatin of chromosomes 1, 9, 15, 16 and the distal long arm of Y
Telomere banding	Bands the terminal regions of chromosomes. The method is similar to R-banding but is more destructive. The bands produced are at the most distal terminal portions of the chromosomes. It is useful for studying rearrangements involving the telomeres of chromosomes which may not be detected with G-banding
G-11 banding	The Giemsa stain is at pH 11 instead of pH 6.8 as for G-banding. Used in human–mouse somatic cell hybrids (Chapter 14) to distinguish mouse and human chromosomes. Human chromosomes stain blue with magenta centromeres and mouse chromosomes stain uniformly magenta with blue centromeres
Kinetochores staining	There are a number of ways to stain kinetochores, including fixation, and ageing regimes, or immunofluorescence using antibodies from scleroderma patients. Detects the point of attachment of chromosomes to the mitotic spindle. The technique identifies pairs of dots at the chromosome centromeres, which may represent the kinetochore or associated chromatin. It is used for investigating kinetochore inactivation in dicentrics
Restriction endonuclease G-banding	Pretreatment with a restriction endonuclease before staining with Giemsa. Used for rapidly identifying unusual polymorphisms

Fig. 7.2). Protocols 22, 23 and 24 describe quinacrine banding (Q-banding) (Fig. 7.3), reverse banding (R-banding) with acridine orange (Fig. 7.4), and constitutive heterochromatin banding (C-banding) (Fig. 7.5), respectively. Other protocols can be found in ref. 2.

G- and R-banding patterns can be produced by staining other than with Giemsa or Leishman's stain. For example, in chromosome painting procedures, a mixture of DAPI and propidium iodide in the mountant for the chromosomes will produce a G-banding pattern when viewed under UV filters and an R-banding pattern when viewed under the green filter set (see Chapter 10, Protocol 53), which can be used to identify individual chromosomes. Also, the chromosome-specific paints obtained by Alu-PCR using Alu primers generate reproducible R-type banding patterns when hybridized to metaphase chromosomes (Chapter 10).

7.4 Detection of constitutional abnormalities

Cytogenetic analysis is generally performed only on

selected patients and their families, and not as a general screening method, because the tests are labour intensive and expensive. Cytogenetic analysis may be carried out on women suffering recurrent abortions and their partners, and on patients with abnormal phenotypes. The individual's constitutional karyotype is established at fertilization; hence if chromosomal abnormalities are present, development may be impaired. The most common abnormalities observed are trisomies and balanced translocations that can be carried through generations. The translocations are a likely cause of recurrent abortions. Unbalanced translocations usually result in fetal loss. Described below are the referral categories used in detection of chromosomal abnormalities.

First 12-week period of gestation Most abnormal conceptions are lost in the first 12-week period. Trisomies of any variety should be expected, some monosomies, unbalanced translocations and triploids.

12 weeks to term Only certain trisomies reach the



Fig. 7.2 Cell stained by NOR banding (see Table 7.2). Acrocentric chromosomes, some of which show positive NOR staining at the nucleolar organizer regions. This patient has an unbalanced karyotype

48,XY,der(11)t(11;13)(q23;p12)mat. The der(11) chromosome shows positive NOR staining at the end of the long arm (arrow). From [13], by permission of Oxford University Press and Dr P.A. Benn.



Fig. 7.3 Quinacrine-banded cell (top) with karyotype (below) from a normal male, 45, XY. See Protocol 22. From [13], by permission of Oxford University Press and Dr P.A. Benn.

stage of 12 weeks to term, including 13, 18, and 21. Unbalanced Robertsonian translocations producing the trisomies 13 and 21, 45,X and other causes of Turner's syndrome, and triploids may be found.

Neonates Certain congenital abnormalities in neonates are associated with cytogenetic abnormalities,

and the cytogeneticist should be familiar with these. As in the earlier categories, trisomies 13, 18 and 21 are expected, unbalanced Robertsonian translocations and 45,X. In addition, *de novo* deletions of 4p, 5p, 9p, 13q, 18 p and q, ring chromosomes, and gross or subtle unbalanced structural rearrangements are found.

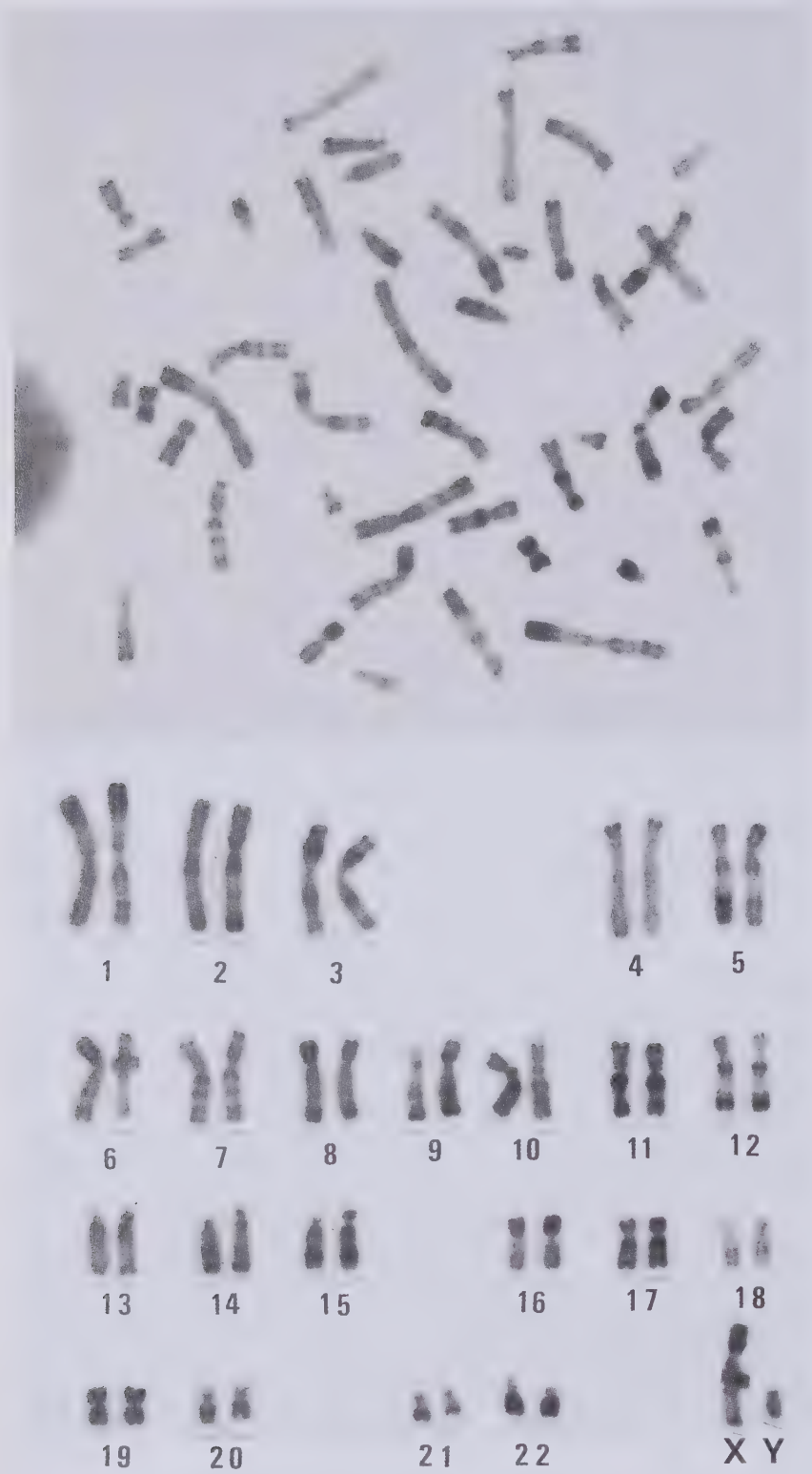


Fig. 7.4 Reverse-banded cell (top) with karyotype (below) from a normal male, 46,XY. See Protocol 23. From [13], by permission of Oxford University Press and Dr P.A. Benn.

Children If the individual has reached childhood, but persists in failure to achieve the developmental milestones, less obvious congenital abnormalities may be present. When karyotyped, small familial or *de novo* rearrangements, interstitial deletions,

marker chromosomes, fragile X and mosaics may be observed. With improved techniques, microdeletions are proving to be more important than previously realized, and may account for defects previously thought to be at a molecular level.



Fig. 7.5 C-banded (constitutive heterochromatin banded) cell from a phenotypically normal male with an inverted Y chromosome; 46,XY,inv(Y)(p11q11). The Y chromosome is marked by the arrow. From [13], by permission of Oxford University Press and Dr P.A. Benn.

Puberty Problems with the sex chromosomes are normally only realized when the individual fails to develop correctly during puberty. The following abnormalities may be found: 45,X, deleted and rearranged X chromosomes (i.e. rings), XY females, XX and XXY males.

Infertility Infertility is caused by a wide range of problems, some of which can be investigated cytogenetically. All the sex chromosome abnormalities noted above can be expected; in addition, XYY, balanced rearrangements, marker chromosomes, X or Y autosome translocations (very rare) and Y structural rearrangements have been associated with infertility.

7.4.1 Approaches to analysis

The average number of cells analysed per sample is usually five, with about six cells counted to eliminate certain degrees of mosaicism. Counting 10 cells excludes 21% mosaicism, 26% mosaicism and 37% mosaicism at 90%, 95% and 99% confidence levels, respectively. 'Counting' refers to just checking the exact number of chromosomes present, whereas 'analysis' identifies each chromosome individually, using the banding pattern to recognize and check all areas of each chromosome. Each laboratory will have different criteria regarding numbers of cells counted and analysed; however, as a guide, use the following:

1 0–12 week/termination, neonatal deaths, stillbirths, abnormal ultrasound (using fetal tissue, villi,

skin, placenta or amniotic fluid): Five cells analysed and five cells counted.

2 Neonates with abnormality suspected on clinical grounds (using blood): Five cells analysed, 25 counted.

3 Neonates or children with dysmorphic features and/or developmental delay, adults with infertility, fetal loss, and others (using blood): Five cells analysed, five counted.

4 Fragile X (blood): Five cells analysed, 100 counted and scored for fragile site at Xq27.3.

5 Sex chromosome abnormality suspected clinically (blood): Five cells analysed, 25 counted.

6 Malignancy investigations (marrow and/or blood): Fifteen to 20 cells analysed (see also Section 7.5.2).

If an odd cell is found in a complete analysis it is only common sense to scan a further 20–50 cells for the anomaly. Each case is usually treated with an individual approach, depending on the findings. A code of practice for clinical cytogenetics has been produced in the USA [4]. Experience will eliminate the need to assess normal variants by various chromosome identification techniques including C-banding, but uncertain cases can be proved to be normal or otherwise by using extra techniques. Abnormality rates vary within the different types of referral category. In general, the abnormality rate will not be greater than 15% overall. In categories of patients with phenotypic abnormalities, however, the rate is higher. Table 7.3 shows a range of chromosomal abnormalities and their main associated clinical features. Table 7.4 provides a guide to the conventions of cytogenetic nomenclature.

7.5 Cancer cytogenetics

When cytogenetic analysis by G- and C-banding techniques was first applied to tumour cells in the 1970s, it became apparent that many malignant and premalignant cells had acquired chromosomal abnormalities. As more samples were studied, it became clear that these abnormalities did not occur at random. Some were highly specific to particular diseases (e.g. the translocation t(15;17) in acute promyelocytic leukaemia, APLM; see Fig. 7.8); some were less specific (e.g. trisomy 8 in a range of myeloid disorders). Cytogenetic analysis is now a crucially important part of the diagnosis and clinical management of cancer patients, and has greatly increased our understanding of tumorigenesis. The consistent cytogenetic changes in some malignancies have given clues to the whereabouts of genes involved in tumorigenesis (Table 7.5). The combination of cytogenetics, molecular genetic

Table 7.3 Chromosomal changes in genetic disorders.

Abnormality	Syndrome	Clinical features
del(3)(q26)	Cornelia de Lange	Growth and mental retardation, arched eyebrows, upper limb defects
del(4)(p16.3)	Wolf–Hirschhorn	Growth and mental retardation, cleft lip, ‘Greek helmet’ nasal bridge
del(5)(p15)	Cri-du-chat	Kitten-like cry, retardation, microcephaly
trisomy 8 (mosaic)		Large square skull, broad nose, long thorax, slender body
del(8)(q24.11q24.13)	Langer–Giedion	Bulbous nose, sparse hair, microcephaly
del(11)(p13)	WAGR	Wilm’s tumour, aniridia, genitourinary abnormalities, mental retardation
dup(11)(p15.5)	Beckwith–Wiedemann	Overgrowth, macroglossia, exomphalos
trisomy 13	Patau’s	Blindness, deafness, epilepsy, cleft lip, polydactyly
del(13)(q14.2)	Retinoblastoma	Retinoblastoma, osteosarcoma, bossed forehead
del(15)(q11q13)	Angelman	Severe retardation, jerky gait, epilepsy
del(15)(q11q13)	Prader–Willi	Hypotonicity, hypogonadism, obesity
del(17)(q11.2)	Neurofibromatosis	café-au-lait patches, neurofibromata
trisomy 18	Edward’s	Growth retardation, rockerbottom feet, brain, heart and gut malformations
del(20)(p11.23p12.1)	Alagille	Pulmonary artery stenosis, biliary hypoplasia, deep-set eyes, large forehead
trisomy 21	Down’s	Flattened facial profile, epicanthic folds, brachycephaly
del(22)(q11.21q11.23)	Cat-eye	Colomata, ear tags, heart defects
45,X	Turner	Short stature, webbed neck, primary amenorrhoea
47,XXX	Triple X female	Occasional delay in mental development
47,XXY	Klinefelter’s	Gynaecomastia, eunuchoid features, advanced growth
47,YYY		?Violent temperament
69,XXX/69,XXY	Triploidy	Mental retardation

analysis, immunological analysis, and observation of cellular morphology can now provide a comprehensive picture of the condition, confirming a diagnosis, and giving disease type, status and prognostic information. Patient management has been improved by the close liaison between cytogeneticist, immunologist and clinician in dealing with these cases.

For example, cytogenetic analysis can help evaluate whether patients with chronic granulocytic leukaemia (CGL)/chronic myeloid leukaemia (CML) are entering a blast crisis. When they are followed up cytogenetically, patients with CGL/CML show further abnormalities in addition to t(9;22) (which gives rise to the Philadelphia chromosome), which are considered to be an indication of transformation to acute leukaemia or blastic crisis. The time until blastic crisis occurs cannot be predicted accurately: if extra abnormalities are seen it is not always the case that blastic crisis will occur immediately. Occasionally, the abnormalities precede blastic crisis by several months. Table 7.6 shows the additional changes commonly found in transformed CGL/CML and the frequency with which they

occur. The additional changes can appear alone or in association with each other; for example, +8 and i(17q) can be observed together or in any other variation, or all the changes can be observed simultaneously, as in +8, i(17q), +der(22) and +19.

The following sections summarize the type of sample required, the culture conditions, and approaches to chromosome analysis of cancer cells in haematological malignancies. Tables IX.1–IX.10 in Appendix IX give details of the classifications of haematological malignancies and a list of abnormalities known to be associated with acute myeloid leukaemia (AML), acute lymphoid/lymphocytic leukaemia (ALL), myelodysplastic syndromes (MDS), myeloproliferative disorders (MPD), lymphomas, and chronic lymphoproliferative disorders. Figures 7.6–7.9 show some cytogenetic abnormalities associated with a range of haematological malignancies. The cytogenetic analysis of solid tumours is covered in Chapter 8, and tables of chromosomal abnormalities found in a range of solid tumours are also given in Appendix IX. Many other changes are seen in malignancies, and indeed, new associations are

Table 7.4 Conventions of cytogenetic nomenclature.

A–G	chromosome groups
ace	acentric fragment
add	additional material of unknown origin
< >	to indicate ploidy level, e.g. <2n>
→	‘from/to’ when describing derivative chromosomes
b	break
c	constitutional karyotype
cen	centromere
chi	chimaera
:	break (in detailed descriptions)
::	break and reunion
, (comma)	separates chromosome numbers, sex chromosomes and chromosome abnormalities
del	deletion
der	derivative chromosome
dic	dicentric
dmin	double minutes
dup	duplication
f	fragment
fra	fragile site
hsr	homogeneously staining region
i	isochromosome
ider	isoderivative
idic	isodicentric
inc	incomplete karyotype
ins	insertion
inv	inversion
mar	marker chromosome
min	minute
– (minus)	chromosome loss
p	short arm of chromosome
p10	short arm part of the centromere
()	surrounds structurally altered chromosomes
Ph ¹	Philadelphia chromosome
+ (plus)	chromosome gain
prx	proximal
q	long arm of chromosome
q10	long arm part of centromere
?	query in identification of chromosome
r	ring chromosome
s	satellite
sce	sister chromatid exchange
; (semicolon)	separates chromosomes and regions in structural rearrangements involving more than one chromosome
/	separates cell lines in mosaics and chimaeras
t	translocation
tas	telomeric association
ter	terminal region of chromosome
tr	triradial
trc	tricentric
trp	triplication

constantly being discovered, so the cytogeneticist should be aware of anything and everything when analysing these types of cases!

7.5.1 Practical considerations

The type of sample being analysed depends on the disease that is being studied. Bone marrow is the

tissue of choice for haematological disorders; peripheral blood in addition to bone marrow is useful in chronic disorders such as CML and chronic lymphocytic leukaemia (CLL). It is important to use lymph node tissue in order to study the chromosome changes in lymphoma, as the bone marrow is not always involved in disease if diagnosed early therefore a lymph node biopsy is required.

Table 7.5 Genes involved in translocations and inversions in haematological malignancies.

Abnormality	Respective genes	
<i>Acute myeloid leukaemia/acute non-lymphocytic leukaemia (AML/ANLL)</i>		
inv(3)(q21q26)		EV11
t(3;3)(q21;q26)		EV11
t(6;9)(p23;q34)	DEK	CAN
t(8;21)(q22;q22)	ETO	AML1
t(9;11)(p22;q23)	AF9/MLLT3	MLL
t(11;19)(q23;p13)	MLL	ENL
t(15;17)(q22;q21)	PML	RARA
<i>Acute lymphocytic leukaemia (ALL)</i>		
<i>Pre-B cell and B cell</i>		
t(1;19)(q23;p13)	PBX1	E2A
t(2;8)(p12;q24)	IGK	MYC
t(8;14)(q24;q32)	MYC	IGH
t(8;22)(q24;q11)	MYC	IGL
t(17;19)(q22;p13)	HLF	E2A
<i>Mixed</i>		
t(4;11)(q21;q23)	AF4	MLL
t(9;22)(q34;q11)	ABL	BCR
t(11;19)(q23;p13)	MLL	E2A
<i>T cell</i>		
t(1;7)(p34;q34)	LCK	TCRB
t(1;14)(p32;q11)	TCL5	TCRD
t(2;8)(q24;q24)	TCL4	MYC
t(7;2)(q34–35;q34)	TCRB	TCL4
t(7;19)(q34–36;p23)	TCRB	LYL1
t(8;14)(q24;q11)	MYC	TCRA
t(10;14)(q24;q11)	HOX11(TCL3)	TCRD
t(11;14)(p13;q11)	TCL2	TCRD
t(11;14)(p15;q11)	TCL1	TCRD
inv(14)(q11q32)	TCRA	IGH
<i>Chronic lymphocytic leukaemia (CLL)</i>		
<i>B cell</i>		
t(2;14)(q13;q32)	REL	IGH
t(14;19)(q32;q13)	IGH	BCL3
<i>T cell</i>		
t(8;14)(q24;q11)	MYC	TCRA
inv(14)(q11q32)	TCRA	IGH
<i>Multiple myeloma</i>		
t(11;14)(q13;q32)	BCL1	IGH
<i>Chronic myeloid leukaemia/chronic granulocytic leukaemia (CML/CGL)</i>		
t(9;22)(q34;q11)	ABL	BCR
<i>Non-Hodgkin's lymphoma</i>		
t(2;8)(p12;q24)	IGK	MYC
t(8;14)(q24;q32)	MYC	IGH
t(8;22)(q24;q11)	MYC	IGL
t(11;14)(q13;q32)	BCL1	IGH
t(14;18)(q32;q21)	IGH	BCL2

In B or T cells the translocation either brings a potential oncogene (e.g. *MYC*) under the control of an active immunoglobulin or T-cell receptor gene (*IGH*, *IGK*, *IGL* and *TCRA*, *TCRA*, *TCRG*, *TCRD* respectively) or creates a fusion gene that encodes a novel fusion protein (e.g. *ABL-BCR*). Many of the genes involved encode known transcription factors.

Table 7.6 Additional chromosome changes found in transformed CGL/CML.

Change	Frequency (%)
trisomy 8	60
i(17q) and abnormalities of 17p	50
+der(22) Philadelphia chromosome	40
trisomy 19	20

The methods of processing the samples are similar to those already given in Protocol 13, but care must be taken in setting up the samples for culture because of the variability in white blood cell count. Some laboratories actually count the number of cells in order to give an optimum cell number of 10^6 per ml per culture. However, experience in handling these samples can allow this step to be eliminated. Generally, two or three 5- to 10-ml cultures can be set up from about 1 ml of bone marrow. The samples are taken into universal containers with 5 ml of transport medium (Table 7.7).



Fig. 7.6 $t(6;9)(p23;q34)$. G-banded chromosomes. This is a very rare and subtle translocation, difficult to detect on poor preparations. It is found in relatively young AML patients.



Fig. 7.8 $t(15;17)(q22;q21)$. G-banded chromosomes. This translocation is found exclusively in APML.



Fig. 7.7 $t(9;22)(q34;q11)$. G-banded chromosomes. This translocation is associated with chronic phase CGL/CML.

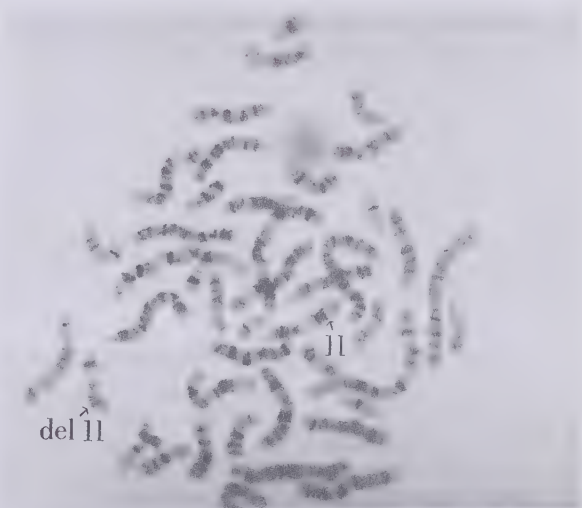


Fig. 7.9 $del(11)(q21q25)$. G-banded chromosomes. This deletion is observed in MDS.

Table 7.7 Bone marrow medium.

McCoys 5A or RPMI	100 ml
FCS ^a	20 ml
Penicillin (10 000 IU ml ⁻¹)	1 ml
Streptomycin (10 000 µg ml ⁻¹)	1 ml
L-glutamine (200 mM)	1 ml

^aFor transport medium replace FCS with 1 ml of 1000 IU ml⁻¹ preservative-free heparin.

The samples are centrifuged at 1000 r.p.m. for 10 min when they arrive in the laboratory. The transport medium is removed and the sample is set up in culture medium. The type of cultures generally set up are: ‘direct’, ‘overnight’, ‘overnight with colcemid’ (0.1 ml colcemid (0.02 µg ml⁻¹)), and a synchronized culture. This laboratory routinely performs overnight with colcemid and synchronized cultures. Care should be taken when dealing with cells from patients with chronic disorders such

as CML and CLL, as there tend to be a large number of active bone marrow cells which will overgrow in culture (and occasionally in transit), and produce no divisions; hence only a small amount of marrow should be used per culture. Once set up, the sample can be processed as in Protocol 13 from step 3 onwards, except that it should be incubated in KCl for 15 min (add colcemid the next day for 1–3 h if *not* performing ‘overnight with colcemid’ culture). This method can be used for blood and marrow samples. The synchronization method for bone marrows is slightly different from that given in Protocol 14 for lymphocytes, so Protocol 25 should be used. When setting up samples with a large number of leukocytes, such as those from patients with CML or CLL, it can sometimes be advantageous to separate the white cells prior to setting up the culture (see Chapter 11, Protocol 55).

Occasionally, only symptoms will appear on request forms, without any proposed or suspected diagnoses. Table 7.8 lists the various symptoms

Table 7.8 Haematological diseases and their symptoms.

Disease and haematological findings	Symptoms associated with the disease
<p><i>Acute leukaemia</i></p> <p>Normochromic, normocytic anaemia</p> <p>White cell count decreased, normal or increased</p> <p>Thrombocytopenia (can be extreme in AML)</p> <p>Variable numbers of blast cells in blood film. AML films may contain Auer rods, and other abnormal cells may be present: promyelocytes, agranular neutrophils, myelomonocytic cells</p> <p>Hypercellular bone marrow, marked proliferation of blast cells, typically over 75% of the marrow cell total</p> <p>Disseminated intravascular coagulation in AML M3</p>	<p>Due to marrow failure:</p> <p>Pallor, lethargy, anaemia</p> <p>Fever, malaise, features of infections, including septicaemia</p> <p>Spontaneous bruises, purpura, bleeding gums and bleeding from venepuncture sites due to thrombocytopenia</p> <p>Due to organ infiltration:</p> <p>Tender bones, especially in children</p> <p>Superficial lymphadenopathy (ALL)</p> <p>Moderate splenomegaly, hepatomegaly (ALL)</p> <p>Gum hypertrophy and infiltration, rectal ulceration, skin involvement (particularly AML M4 and M5)</p> <p>Meningeal syndrome (ALL), headache, nausea and vomiting</p> <p>Testicular swelling (ALL)</p> <p>Mediastinal compression (particularly T-cell ALL or T lymphoblastic lymphoma)</p>
<p><i>Chronic granulocytic leukaemia (CGL)/chronic myeloid leukaemia (CML)</i></p> <p>Leukocytosis usually $>50 \times 10^9 \text{ l}^{-1}$ and up to $500 \times 10^9 \text{ l}^{-1}$</p> <p>Complete spectrum of myeloid cells in peripheral blood</p> <p>The levels of neutrophils and myelocytes exceed those of blast cells and promyelocytes</p> <p>Hypergranular marrow with granulopoietic predominance</p> <p>Increased circulating basophils</p> <p>Platelet count normal, decreased or increased</p> <p>Neutrophil alkaline phosphatase (NAP) score low</p>	<p>Hypermetabolism, e.g. weight loss, lassitude, anorexia, and night sweats</p> <p>Splenomegaly nearly always present and sometimes massive. The enlargement can cause discomfort, pain or indigestion</p> <p>Pallor, dyspnoea and tachycardia, anaemic features</p> <p>Bruising, epistaxis, menorrhagia or haemorrhage from other sites</p>

Continued.

Table 7.8 Continued.

Disease and haematological findings	Symptoms associated with the disease
<i>Chronic lymphocytic leukaemia (CLL)</i> Leukocytosis between 5 and $300 \times 10^9 \text{ l}^{-1}$; 70–90% of white cells on blood film appear as small lymphocytes Normocytic, normochromic anaemia Thrombocytopenia Bone marrow shows lymphocytes comprising 25–95% of all cells Reduced concentration of serum immunoglobulins	Symmetrical enlargement of superficial lymph nodes Pallor, dyspnoea Splenomegaly and hepatomegaly Bruising in patients with thrombocytopenia Pruritus associated with herpes zoster virus Tonsillar enlargement
<i>Hairy cell leukaemia (HCL)</i> The ‘hairy cells’ (a type of B lymphocyte) are present in blood, liver and other organs Trepphine shows mild fibrosis Serum paraprotein may be present	Spleen moderately enlarged Pancytopenia Disease peak at 40–60 years of age with a male to female ratio of 4 : 1
<i>Myelodysplastic syndromes (MDS)</i> Qualitative and quantitative abnormalities in one or more of the three myeloid cell lines: red cells, granulocytes and monocytes and platelets Wide range of abnormalities in peripheral blood and bone marrow: macrocytosis, ring sideroblasts, megaloblastic erythropoiesis, disordered granulopoiesis and megakaryocytes	Anaemia Infections due to impaired phagocytic production and/or function
<i>Hodgkin’s disease (HD)</i> Normochromic, normocytic anaemia, with marrow failure and infiltration Leukocytosis in one-third of patients due to increased numbers of neutrophils Neutrophil alkaline phosphatase (NAP) score increased Eosinophilia Lymphopenia (advanced disease) Platelet count normal or increased in early disease but low in later stages Erythrocyte sedimentation rate (ESR) raised Bone marrow involvement rare in early disease	Painless, non-tender, asymmetrical, firm, discrete enlargement of superficial lymph nodes Splenomegaly in 50% of patients. The liver may be enlarged Mediastinal involvement in 6–11% patients (nodular sclerosis type in women) Cutaneous Hodgkin’s disease occurs as a late complication Also seen: fever, pruritus, alcohol-induced pain, weight loss, profuse night sweats, weakness and fatigue
<i>Non-Hodgkin’s lymphoma (NHL)</i> Normochromic, normocytic anaemia; also autoimmune haemolytic anaemia may develop When bone marrow is involved, neutropenia, thrombocytopenia or leukoerythroblastic features Lymphoma cells may be present in peripheral blood Trepphine shows focal involvement in about 20% cases Diffuse infiltration and fibrosis may occur	Median presentation age, 50 years Superficial lymphadenopathy Fever, night sweats and weight loss less frequent than in HD and usually indicate disseminated disease. Anaemia and infections Oropharyngeal involvement, sore throat, obstructed breathing in 5–10% of patients Abdominal disease, liver and spleen often enlarged Skin, brain and testis or thyroid involvement. The skin is also primarily involved in two closely related T-cell lymphomas: mycosis fungoides and Sézary’s syndrome
<i>Burkitt’s lymphoma (BL)</i> B-cell lymphoblastic lymphoma Isolated histiocytes in masses of abnormal lymphocytes produce the ‘starry sky’ appearance in tissue sections Epstein–Barr virus identified in Burkitt cell culture	Predominantly in young African children Massive jaw lesions Extranodal abdominal involvement Ovarian tumours (in girls) Severe pruritus and psoriaform lesions Lymph nodes, spleen, liver and bone marrow ultimately affected
<i>Mycosis fungoides and Sézary’s syndrome</i> Circulating T lymphocytes Cutaneous T-cell lymphoma	Exfoliative dermatitis Erythroderma Generalized lymphadenopathy

Table 7.8 Continued.

Disease and haematological findings	Symptoms associated with the disease
<i>Multiple myeloma</i> In 98% of patients monoclonal protein occurs in serum and/or urine Bence-Jones protein occurs in two-thirds of cases Bone marrow shows increased plasma cells Normochromic, normocytic or macrocytic anaemia Rouleaux formation in red cells Neutropenia and thrombocytopenia in advanced cases Peripheral blood film shows abnormal plasma cells (15% patients) Serum calcium elevation (45%) Blood urea raised (20%) Low serum albumin in advanced disease	Bone pain (especially backache) Lethargy, weakness, dyspnoea, pallor, tachycardia due to anaemia Repeated infections caused by deficient antibody production and later due to neutropenia Anorexia, vomiting, constipation and mental disturbance due to renal failure Abnormal bleeding tendency: myeloma proteins interfere with platelet function and coagulation factors
<i>Waldenström's macroglobulinaemia</i> Seen mostly in males over 50 years of age Proliferation of cells which produce monoclonal IgM paraprotein Blood viscosity increased High ESR Peripheral blood lymphocytosis Bone marrow infiltration by small lymphocytes, plasma cells, 'plasmacytoid' forms, immature lymphoid cells, mast cells and histiocytes	Fatigue and weight loss Hyperviscosity syndrome Engorged veins in retina Bleeding tendency Anaemia due to haemodilution, decreased red cell survival, blood loss, bone marrow failure Moderate lymphadenopathy, enlargement of liver and spleen
<i>Polycythaemia rubra vera (PRV)</i> Haemoglobin, haematocrit and red cell count increased Neutrophil leukocytosis (in over half of patients) Raised platelet count (in half of patients) NAP score increased Increased serum vitamin B ₁₂ binding capacity Hypercellular bone marrow with prominent megakaryocytes Blood viscosity increased	Headaches, pruritus, dyspnoea, blurred vision and night sweats Retinal venous engorgement, conjunctival suffusion Splenomegaly (in two-thirds of patients) Haemorrhage or thrombosis Gout
<i>Essential thrombocythaemia (ET)</i> Abnormal large platelets and megakaryocyte fragments in peripheral blood film Platelet count raised above 1000 Platelet function tests abnormal	Anaemia Massive splenomegaly giving discomfort, pain or indigestion Weight loss, anorexia and night sweats Bleeding problems and bone pain

Adapted from ref. 11.

associated with haematological malignancies. These can give some indication of the possible disease involved and thus aid in setting up samples.

For lymph nodes or marrows from lymphoma patients, the cultures require stimulation with a B-cell mitogen (the vast majority of lymphoproliferative disorders involve B cells). This laboratory normally prepares an overnight with colcemid culture and then 3-day cultures with and without the mitogen 12-*O*-tetradecanoylphorbol-13-acetate (TPA). Make up a stock solution of 100 µg ml⁻¹ TPA and use a final concentration of 50 ng ml⁻¹. For known T-cell disorders, a cocktail of mitogens

should be used, including pokeweed mitogen (PWM), PHA and TPA. Colcemid can be added on the second day, last thing at night and processed the next morning. If a lymph node is provided by the clinician, this needs to be macerated before culture, and the culture method is similar to that used for chorionic villi (see Protocol 18). However, as the culture is in suspension, and short term, there is no need to adhere the cut pieces to the culture vessel; alternatively, the pipette is used to push the pieces of tissue through a gauze to extract the larger fragments.

7.5.2 Approaches to analysis

Care is required when analysing samples from patients with malignancy. It is often the case that the abnormal clone is concealed in the poorer-quality cells. It is very tempting to analyse only the good-quality metaphases, but during training, the cytogeneticist must be taught to examine poor quality metaphases as well. Occasionally, only certain types of culture display the abnormal clone; for example, the t(15;17) in APL is *not* seen in direct cultures. It is always advisable to analyse some cells from all the cultures that have been set up. Normally, this laboratory fully analyses 15–20 cells per referred sample. This is sufficient to detect the large majority of mosaics. If enough cells are studied, there will invariably be some normal cells in addition to the abnormal cells. The detection of no abnormalities does not necessarily indicate that the patient does not have an abnormal clone. It is possible that: (i) the abnormality is so subtle that it is beyond the limit of the microscope, or (ii) the abnormal line is a very small percentage of the cells, or (iii) the abnormal line has not been stimulated sufficiently in culture, or those cells have not been captured in mitosis. Once an abnormality has been established, assessing a patient in the future becomes easier, and full analysis need not be carried out on every single cell. It is not prudent to report on one abnormal cell and indeed, the definition of clonal abnormality is described in detail in *Guidelines for Cancer Cytogenetics* [5] and is as follows:

- two or more cells have the same structural abnormality;
- two or more cells have acquired the same chromosome (trisomy);
- three or more cells have lost the same chromosome (monosomy).

When an abnormality found in a single cell cannot be observed in further cells, it should be ignored. The only exception to this would be where a specific abnormality is found that would concur with the clinical diagnosis, for example the t(9;22) translocation in a sample that was suspected of having CGL, or the t(15;17) in a suspected case of APL, or other such cases where the association with a particular disease type is strong. Generally, common sense should be exercised in reporting such cases, and a close liaison with the clinician in charge is always an advantage.

Due to the fact that the abnormal cells are often of poorer quality, it is very important that the cytogeneticist has gained enough experience in other fields, such as prenatal work, to enable an objective approach to be applied to analysis. It can

be very difficult to assess whether a small deletion is present or whether the 'abnormality' is caused by a cultural artefact. Many frustrating hours can be spent in decisions of this nature, knowing the implications to prognosis that a certain result may carry. Experience in observing various translocations and their abnormal chromosome products can also be extremely useful when preparations are poor, as once a translocation product has been observed, it tends to become imprinted in the mind and can be recognized amidst a plethora of other changes if required! There is no substitute for hands-on experience of these kind of samples.

7.6 Digital imaging

The detection and analysis of chromosomal abnormalities is very labour intensive and can therefore prove expensive. There are certain aspects of this part of the cytogeneticist's job which can be aided by automated imaging and detection systems using digital imaging (see Chapter 13 for description of the technology as applied to fluorescent imaging). The currently available systems have four main applications: metaphase finding, karyotyping, image enhancement and presentation.

Although manual techniques are normally sufficient for metaphase selection, in direct chorionic villus cultures, fragile X preparations, and poor bone marrow preparations, it can prove time consuming to locate sufficient metaphases for analysis in preparations with a low mitotic index. Here, an automated system can prove useful, and some systems can be set up for overnight scanning. For metaphase selection it is important that the automated system should be able to distinguish between cellular debris and any quality of metaphase.

For karyotyping, an imaging system should be able to deal with overlapping chromosomes and be able to rotate the image of a chromosome to its conventional axis, but only the most expensive systems can do this. It is important to assess the capabilities of a system before purchase, as needs of different laboratories vary. Ease of use is of course the most important factor.

When digital fluorescent imaging using a highly sensitive camera and computerized enhancement can be applied (see Chapter 13), the image following enhancement is superior to that obtained by darkroom methods. Image enhancement alters the intensity levels of the banding patterns electronically to suit the user and can resolve bands that are close together and can improve contrast (see, for example Fig. 13.9).

Teaching and publication benefits from the

presentation facilities available on some imaging systems. Interpretation of cases can be enhanced by the ability of the system to bring the Paris convention diagrams onto the screen next to the karyotype, and also to display several chromosomes from different cells for comparison, but again only the most expensive ones can do this. The chromosomes can also be straightened and enlarged. The effectiveness of a computerized image system is enhanced by spreading the slide thinly, washing the maximum amount of debris from the suspensions and not staining too darkly. Finally, *never* rely solely on what the computer finds!

Automated photography is a labour-saving aspect of the computerized image analysis system. A high-quality laser printer provides the cytogeneticist with prints of photographic quality. This is extremely useful for a permanent record of each case as an alternative to the tedious darkroom methods of the past. Unfortunately, the prints do discolour at present, but eventually this problem may be rectified. These systems are very expensive, and maintenance and depreciation are important factors to be taken into consideration. Also, with technology improving constantly, models soon become obsolete. However, we watch this space with interest!

.....

Protocol 13

Processing of whole blood samples and slide preparation

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- Complete medium: 100 ml Ham’s F10 or RPMI 1640,^a 10 ml fetal calf serum (FCS),^b 1.0 ml phytohaemagglutinin (PHA) (purified), 1.0 ml penicillin (5000 IU ml⁻¹), 1.0 ml streptomycin (5000 µg ml⁻¹), 1.0 ml L-glutamine (200 mM)
- whole blood
- colcemid stock solution (10 µg ml⁻¹)
- 0.075 M KCl
- fixative (methanol:glacial acetic acid, 3:1)
- universal containers or Leighton tubes
- clean glass slides

Table 7.7 gives medium for bone marrow culture.
^aUse RPMI or McCoy’s 5A for bone marrow culture.
^bUse 20 ml FCS for bone marrow culture.

Method

- 1

Use either 10 ml complete medium in a universal container or 5 ml in a Leighton tube. Inoculate with the appropriate amount of whole blood. For a 10-ml culture:

adults and children over 5 years old	0.8 ml
children less than 5 years old	0.5 ml
infants up to 5 years old	0.1 ml
cord blood	0.3 ml
fetal blood	0.2 ml
- 2

Incubate the culture at 37 °C for 48 or 72 h.
- 3

Add 0.1 ml colcemid solution (final concentration, 0.1 µg ml⁻¹) for 2 h before harvesting cells, in order to arrest the mitoses at metaphase.

- 4 Transfer to a centrifuge tube^a and spin at 1000 r.p.m. for 5–10 min.
- 5 Remove the supernatant, mix the pellet thoroughly, then resuspend in either 5 or 10 ml 0.075 M KCl depending on the original culture size (the KCl should be warmed to 37 °C). Incubate at 37 °C for 10 min.^b
- 6 Centrifuge at 1000 r.p.m. for 5–10 min.
- 7 Remove the supernatant, mix the pellet thoroughly, and then add a few drops of chilled fixative. Add 5–10 ml of fresh chilled fixative.
- 8 Repeat steps 6 and 7 until the supernatant is clear (usually two more washes is sufficient).
- 9 On the final centrifugation, remove the supernatant and resuspend the resulting pellet in a small volume of fixative in order to give a milky white suspension. Drop the suspension either onto cold wet slides or clean dry slides (see below). For bone marrows, the preferred slide is cold and wet. Dry slides can be used for amniotic fluid preparations, or blood samples.^c

^aIf using Leighton tubes, the transfer to centrifuge tubes is unnecessary.

^bIncubate for 15 min for bone marrow cultures.

^cIf preparing chromosomes for microdissection by the thymidine block method (see Protocol 14) prepare chromosomes on coverslips as in Chapter 11, Protocol 58.

This protocol can be varied slightly for different tissues—for example, bone marrows (see Protocol 25)—but the general procedure is the same for all types of culture.

Slide preparation

Several factors must be considered in order to obtain good-quality slide preparations. It is important to add the first few drops of fixative slowly, ensuring that the suspension is thoroughly mixed, in order to prevent clumping. If using the prefix stage as described below, shaking well after the addition of the few drops is sufficient to aid the optimum mix. If not, then one has to add a few drops at a time, mixing between each addition until about 1 ml of suspension is achieved; then you can add larger amounts, but mixing/shaking all the time.

Some laboratories use a prefix stage, which entails the addition of a few drops of fixative following incubation in KCl and before centrifugation. This step appears to aid the subsequent mixing process. One can also chill the cells in the freezer for 30 min (or longer if necessary) after fixation before spreading onto the slides.

The slides also need to be clean, and although slides of good quality can be used straight from the box, it is preferable to clean slides in alcohol before use and store in distilled water in the refrigerator prior to spreading.

Some laboratories favour warmed slides for spreading, but generally, good spreading depends on the ambient temperature in the laboratory. A temperature difference between the suspension and slides is required to facilitate adequate spreading.

If using cold wet slides, spreading can be further enhanced by placing the spread slides immediately on a hotplate at 50 °C. The latter method is particularly useful with preparations that are difficult to spread, such as bone marrows.

Troubleshooting

Slides too crowded, interphases overpowering metaphases

This is a common beginner's problem, and generally can be cured by adding more fixative to dilute the final suspension.

- *Always take away more supernatant than necessary initially, as it is always easier to add more fixative than to recentrifuge and resuspend the pellet.*
- *When adding more fixative to final suspension, mix very gently.*
- *Respread slides.*

Slides very sparse, paucity of interphases and metaphases

Also a common problem, caused either by too dilute a final suspension or, occasionally, if a sample is very poor. If the latter, it is probable that the initial concentration of cells was too low for optimum growth; this can often be the case in samples of myelodysplastic bone marrows.

- *For a dilute sample, add more fixative, recentrifuge and then resuspend the pellet in a smaller amount of fixative.*
- *Respread the slides.*
- *For a poor sample, if no more material remains, a repeat sample may have to be requested.*

Metaphases broken up, otherwise known as 'chromosome soup'

This is caused by overspreading the slides, usually because of the ambient conditions during spreading. If there is a large difference between the temperature of the slides and the ambient temperature, slides spread too readily.

- *If you have been using a hotplate to spread slides—don't, in these conditions.*
- *If using cold slides, allow them to warm up gently prior to dropping suspensions.*
- *Respread the slides.*

Alternatively, if this does not improve the slides, it is possible that KCl treatment during processing was excessive for the material concerned.

- *Reduce the time in KCl.*
- *If you have been placing the tubes in the incubator during incubation with KCl, try leaving them on the bench instead, at room temperature.*
- *Be aware that bone marrows require more time in KCl than amniotic fluids, chorionic villi and blood samples.*

Metaphases too clumped, cytoplasm still visible around metaphase plate

This will make banding slides extremely difficult, and is a far greater problem than that above! It can be caused by insufficient temperature

difference between the slides and hotplate during spreading, or, more usually, insufficient time in KCl, or insufficient warming of the KCl.

- *Always warm KCl to 37 °C before use.*
- *Leave tubes for a longer incubation with KCl.*
- *Put KCl-treated tubes in incubator for the incubation period.*
- *During spreading, ensure slides are cold and the suspension is as cool as possible.*
- *Respread slides and put on hotplate immediately.*

Insufficient mitoses for analysis

This problem may be caused by insufficient incubation time with colcemid, or a slow-growing culture, or poor timing in the synchronized cultures.

- *Increase colcemid treatment time.*
- *Check the timing of synchronized cultures.*
- *In the case of slow-growing cultures, such as peripheral bloods from lymphoproliferative disorders or myelodysplastic syndromes, culture cells for 5 days before colcemid treatment.*

Protocol 14 Thymidine block synchronization method for obtaining prometaphase chromosomes

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- complete medium as in Protocol 13, but containing 2.0 ml PHA
- thymidine (1 g in 67 ml PBS)

Method

- 1 Add 0.4 ml whole blood to 7 ml culture medium and incubate at 37 °C for 48 h.
- 2 Add 0.2 ml thymidine and incubate for approx. 16 h. Thymidine blocks the cell cycle and prevents cell division.
- 3 Transfer the culture to a centrifuge tube and centrifuge at 800 r.p.m. for 10 min.
- 4 Remove supernatant and resuspend in 7 ml warmed PBS.
- 5 Centrifuge at 800 r.p.m. for 10 min.
- 6 Remove supernatant and resuspend the pellet in 7 ml complete medium (which can be prewarmed). Incubate at 37 °C for 4–5 h.

^aIf using this protocol to prepare chromosomes for microdissection, prepare chromosomes on coverslips as described in Chapter 11, Protocol 58.

- 7 For the final 15 min of incubation time, add 0.2 ml colcemid solution (10 µg ml⁻¹).
- 8 Process as for Protocol 13, but use 800 r.p.m. centrifugation speed and incubate with KCl for only 5 min.^a

Troubleshooting

Insufficient or no metaphase spreads

If timings above are adhered to, there should not be any problems with this method. Always check times, and don't cut corners here!

Protocol 15 Fragile X detection

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- medium 1—Low-folate medium: TC199 or FX-1 with 2% serum (usually FCS) and 1% Hepes buffer
- medium 2—Folate antagonist medium: RPMI 1640 or TC199 with 5% serum and 1% Hepes buffer
- medium 3—Thymidine synthesis inhibiting medium: RPMI 1640 with 5% serum, plus 15 mg ml⁻¹ thymidine
To each medium add all other reagents as for complete medium in Protocol 13, except FCS
- methotrexate (MTX), stock solution of 1 mg per 2.2 ml. Add 0.1 ml per 5 ml culture
- thymidine, stock solution of 15 mg ml⁻¹. Use a 2% solution in culture, or as per protocol
- colcemid, solution of 10 µg ml⁻¹. Add 0.1 ml (or 0.5–1%) per 5 ml culture

Method

- 1 Set up blood cultures in at least two of the media listed above.
- 2 For Medium 1, incubate for 72 h and process as Protocol 13.
- 3 For Medium 2, incubate for 48 h, then add 0.1 ml MTX (final concentration, 10⁻⁷ M). Incubate for 18 h, transfer to a centrifuge tube and spin at 1000 r.p.m. for 5 min. Remove the supernatant and

resuspend in fresh medium with 0.4 ml thymidine (final concentration, 0.04 mg ml⁻¹). Incubate for 6 h and add 0.1 ml colcemid for the final 15 min. Process as in Protocol 13.

- 4 For Medium 3, incubate for 46 h then add 0.3 ml thymidine (final concentration, 0.45 mg ml⁻¹) for 16 h (total, 72 h). Process as in Protocol 13.

.....

Protocol 16 Setting up amniotic fluid cultures

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- complete tissue culture medium for amniotic fluid culture: 100 ml Ham’s F10 medium, 20 ml FCS,^a 2 ml Ultrosor G, 1 ml L-glutamine (200 mM), 1 ml penicillin or streptomycin (10 000 IU ml⁻¹ or 10 000 µg ml⁻¹), 0.1 ml nystatin (1000 µg ml⁻¹), 1 ml Hepes buffer^b (1 M)
- Leighton tubes

The choice of medium is normally one of laboratory preference, and others can be experimented with. Chang medium can also be used, as it was originally prepared for rapid amniotic fluid culture growth. It is now commonly used for chorionic villus cultures. It is quite expensive but growth times can be accelerated considerably, and no serum supplementation is required. Chang medium should be reconstituted according to the manufacturer’s instructions, using Chang A for open systems and Chang C for closed systems. Antibiotics should also be added as for the complete medium given above.

Method

- 1 Centrifuge the universals with the amniotic fluid at 500 r.p.m. for 10 min.
- 2 Remove the supernatant, leaving about 0.5 ml fluid above the undisturbed cell pellet.
- 3 Retain the supernatant for alphafetoprotein test and acetylcholinesterase test as required.
- 4 Depending on the pellet size, add 3–5 ml complete medium to the pellet. Transfer equally to two or three Leighton tubes. Flasks can also be used as an alternative, but more medium would be required for this.^a
- 5 Incubate at 37 °C.

^aThis level of serum can be reduced when Ultrosor G is used.
^bIf an open system (i.e. with a CO₂ incubator) is used, this is not necessary.

^a*In situ* cultures can also be set up on coverslips [9]. This is a personal choice, but if cell numbers may be low, this is a good method.

Troubleshooting

Infection occurs in culture vessels

This is usually caused by poor aseptic technique. Also check cultures from other operators. If everyone is experiencing the same problem (even senior staff), then it is possible a fungus may have infiltrated the safety hood or perhaps the incubator.

- *Check all tubes thoroughly.*
- *Revise and assess aseptic technique.*
- *Fumigate hood and/or incubator.*
- *Always keep a logbook of the procedures carried out on all samples—that is, which medium, serum, etc., were used in each tube.*
- *If setting up three tubes, always set one up using a different bottle of medium, so that if the medium is infected, at least there is a chance of rescuing the sample.*
- *If using an open system, check the water jacket in the incubator is not infected.*

Protocol 17 Setting up a chorionic villus sample

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

The sample will normally contain not only villi, but also maternal decidua, and this along with other debris must be extracted from the sample before setting it up. The following initial steps should be taken to ensure an uncontaminated sample.

- 1 Wash the material a few times in transport medium, until the tissue pieces are clearly visible.
- 2 Remove the medium with a sterile pipette and add a small drop of culture medium to prevent the tissue drying out.
- 3 Using a stereo microscope and two long, fine syringe needles, remove any maternal tissue that may be adhering to the villus pieces. The villi should be washed once more and checked thoroughly to ensure no contamination is present. Again, this is a matter of experience, but any doubtful tissue should not be used for culture. NB Care taken at this step will prevent any potential problems in the long run!

Protocol 18 Setting-up a long-term villus culture

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

The villi prepared as in Protocol 17 are then set up for long-term culture (7–10 days) using either a maceration technique as illustrated below or an enzymatic dissociation method [9] (see Chapter 8, Protocol 27). (See Protocol 19 for direct method.)

Materials

- Chang medium (make up as manufacturer’s instructions)
- size 22 scalpel blade
- 1-ml pipette
- Leighton tubes and/or flasks

Method

- 1** Place the clean villi in a sterile Petri dish and add two drops of Chang medium to the villi pieces.
- 2** Using a size 22 scalpel blade, cut up the tissue into very fine pieces. The end result should resemble a fine mush of tiny fragments.
- 3** Add 0.4 ml Chang medium to the pieces and mix together. Using a sterile 1-ml pipette, disperse the fragments evenly over the bottom of a flask or Leighton tube and leave to adhere for 1 h.
- 4** Add 2–3ml (Leighton tube) or 5 ml (flask) medium to the culture vessel — very carefully! so as not to dislodge the fragments.
- 5** Incubate at 37 °C and observe after 5–7 days.
- 6** Change medium and harvest when required as in Protocol 20.

.....
Troubleshooting

Poor growth

This could be caused by a poor sample originally or by the villi pieces not being macerated sufficiently.

- *Ensure villi pieces are cut to a very fine mush.*
- *Try to disperse fragments as evenly as possible.*
- *Be very careful when adding medium following the 1 hour adhesion period.*
- *If the sample looks of poor quality, it may be necessary to set up only one culture, although this can be risky.*

Maternal cell contamination

Almost certainly caused by poor preparation and insufficient cleaning of the villi pieces.

- *Double-check the pieces with a stereo microscope.*
- *Take longer than you think it will take to clean up the villi pieces.*
- *Give an extra wash to ensure cleanliness.*

However, very occasionally, a laboratory mix-up could be to blame and this must be ruled out.

- *Care with labelling slides and tubes is essential!*

.....

Protocol 19 Direct villus culture

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

The method below is based on one reported by Simoni *et al.* [10] and modified by Flori *et al.* [11].

Materials

- 3 ml Hanks' balanced salt solution (HBSS)
- 3 ml unsupplemented medium (e.g. Ham's F10)
- 0.3 ml colcemid solution (10 µg ml⁻¹)
- 3 ml sodium citrate 1% solution
- fixative (methanol : glacial acetic acid, 3 : 1)
- absolute methanol
- 70% methanol
- 50% methanol
- 20% methanol
- deionized water
- 60% acetic acid
- Petri dish
- inverted microscope
- hotplate at 40 °C
- clean slides
- bent pipette

Method

- 1 Clean the villus pieces as described in Protocol 17 and place in a Petri dish with 3 ml HBSS and then transfer to 3 ml medium with no supplements. The following procedure can be done directly, or the villi can be placed in the incubator overnight before addition of colcemid.

- 2 Add 0.3 ml colcemid solution and incubate at 37 °C for about 3 h.
- 3 After incubation, remove the medium using a Pasteur pipette and add 3 ml 1% sodium citrate solution. Leave for 10 min at room temperature.
- 4 Remove the citrate solution and add fresh fixative. Remove the fixative and replace with fresh fixative twice more before processing.
- 5 Have the hydration series below prepared, as the next steps have to be followed in rapid succession:
 - (a) absolute methanol;
 - (b) 70% methanol;
 - (c) 50% methanol;
 - (d) 20% methanol;
 - (e) deionized water.Remove the fixative and add the above in the order given, replacing each solution with the next step. Prepare fresh 60% acetic acid.
- 6 Remove the water from the villi pieces and push them down into the crease of the Petri dish. Add a few drops of the 60% acetic acid and tap the dish gently. Observe under an inverted microscope to assess the extent of cell dissociation. Usually 2–3 min is sufficient.
- 7 Put a clean slide onto a hotplate at 40 °C. Put a drop of the villi suspension onto the slide and drag up and down the slide using a bent pipette, trying to avoid touching the surface of the slide.
- 8 Allow slide to dry on hotplate.

.....

Protocol 20

Harvesting amniotic fluid and chorionic villus cultures

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- 1.0 ml Versene solution (1 : 5000)
- 0.5 ml trypsin (0.25%)
- 2.0 ml FCS : water mixture (1 : 9)
- 2.0 ml serum-free medium
- fixative (as in Protocol 19)
- inverted microscope
- centrifuge tube
- clean slides

Method

- 1 Remove the medium from the cell culture.
- 2 Wash with 1.0 ml Versene solution (1 : 5000).
- 3 Remove the Versene and add 0.5 ml trypsin (0.25%).
- 4 Tap the culture tube vessel in order to dislodge the cells. Assess the ‘damage’ under an inverted microscope; usually after 1 min there are sufficient cells floating freely in the medium.
- 5 Add 2 ml FCS : water mixture (1 : 9) to arrest the trypsin action.
- 6 Transfer the cell suspension to a centrifuge tube and add 2 ml serum-free medium. Incubate for 20 min at 37 °C.^a
- 7 Add a few drops of chilled fixative to the tube.
- 8 Centrifuge at 1000 r.p.m. for 10 min.
- 9 Remove the supernatant and add 2 ml of fresh fixative, and ensure the cell suspension is completely mixed.
- 10 Repeat steps 8 and 9 two or three times.
- 11 Following the final spin, remove the supernatant and resuspend the pellet in a few drops of fixative. (The amount will vary greatly, and is usually a matter of experience!)
- 12 Drop the cell suspension onto clean slides.

^aThe original culture tube can be refilled with medium and kept growing at 37 °C as a back-up culture if required.

Troubleshooting

Loss of material during processing

This can be caused by too many washes with fixative and also a too zealous shaking of the tube on the final resuspension.

- Wash a maximum of three times in fixative.
- Tap the tube very gently to resuspend pellet before spreading.

Cells not floating off quickly during trypsinization

Occasionally a culture can be stubborn in this respect and it is important to remove the cells as quickly as possible to avoid damaging the mitotic cells.

- Put the tube in the incubator while trypsinizing.
- Hit the tube quite hard with the palm of your hand during trypsin treatment.
- Set up in situ cultures (which don’t need trypsinizing).

Mitotic index poor

Possibly caused by poor assessment prior to trypsinizing, or insufficient time in colcemid.

- *Increase colcemid incubation period.*
- *Check that cells are rounded-up before processing.*

.....

Protocol 21 Giemsa banding

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- 40 ml saline solution (8.5 g l⁻¹)
- 1.0–1.5 ml Difco Bactotrypsin
- 50 ml buffer (pH 6.8)
- Giemsa (1 : 10) with buffer, or Leishman’s stain^a (1 : 5) with buffer
- Coplin jars

Method

- 1 Make up trypsin solution using Difco Bactotrypsin.^b Add 1–1.5 ml reconstituted trypsin^b to approximately 40 ml saline (8.5 g l⁻¹)^b and put in a Coplin jar. Set up another Coplin jar with saline solution, and one more with buffer (pH 6.8).^b
- 2 Submerge slide in the Coplin jar with trypsin solution for 20–40 s. (This time will vary with different types of culture and age of the slide, and it is wise to do a test slide before banding precious material!)
- 3 Rinse in the saline solution, then in buffer and place either in a Coplin jar with Giemsa solution (made up 1 : 10 with buffer, pH 6.8), or stain horizontally with Leishman’s stain (1 : 5 with buffer pH 6.8).
- 4 Rinse with buffer and distilled water.
- 5 Slides can be examined ‘under water’ using a coverslip before drying to assess the banding.

Care should be taken with banding, depending upon the material. If slides are examined under water and found to be underbanded, they can be destained and treated again with trypsin (see Troubleshooting). However, if they have already been dried and mounted with DPX, rebanding can prove difficult.

^aThree grams Leishman’s powder dissolved in 1 l methanol. Leave on hotplate overnight and then filter in the morning into universal containers. Fill the universals to the maximum level and screw the caps on tight. Keep in dark if possible.

^bWhen using G-banding for microdissection purposes, the saline solution and buffer should be autoclaved and the Leishman’s stain and trypsin solution filtered through a Millipore filter.

Troubleshooting

Chromosomes darkly stained, no real visible bands

Caused by insufficient time in trypsin, or by too concentrated a stain, or by too long a staining time.

- *Decrease concentration of stain.*
- *Decrease time of staining.*
- *Increase the time of trypsin treatment.*
- *Slides can be destained and rebanded in this instance.*

Chromosomes very fuzzy and puffed-up

Usually a problem with bone marrows, but can also be a result of overtreatment with trypsin.

- *Always solid stain and then destain bone marrow preparations before banding.*
- *Decrease trypsin time.*
- *Slides cannot normally be rescued in this instance.*

Chromosomes puffy, pale, ghost-like, bands indistinct

This is almost always caused by too long a trypsin treatment. In contrast to underbanding, an overtreated slide cannot be rescued. It is always safer to underband slides as they can be destained and retrypsinized and/or restained. Occasionally, it may be that indistinct bands are just caused by a poor stain concentration, or insufficient time in stain.

- *Prepare fresh slides, leave to age.*
- *Use a shorter trypsin time.*
- *Check stain concentration is correct.*
- *Check stain has not expired; this is a common problem with Leishman's stain, which is affected by exposure to light and air.*

Always trypsinize a test slide before staining a batch of slides. Slides do vary between cultures and patients. If care is taken, no precious material should be lost. The simplest way to ensure decent preparations is to check each culture and/or patient, by placing a coverslip over a freshly banded slide (mounted in water) and checking the banding with a microscope. If the slide is satisfactory, it is then safe to mount it in DPX in the knowledge that the banding is of optimum quality.

Protocol 22 Quinacrine banding

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Used as an alternative to G-banding in some laboratories, this method, which uses the fluorescent dye quinacrine, can also be used to

identify chromosomes, with a banding pattern resembling G-banding (Fig. 7.3). However, analysis has to be performed using a fluorescence microscope and photography must be carried out before the image fades. The advantage is that slides can be banded instantly without the need for ageing (although the method can also be used on older slides). The following method is based on that of Sumner [12].

Materials

- quinacrine hydrochloride, 0.5 g per 100 ml distilled water
- MacIlvaine’s buffer, pH 5.6: 0.1 M anhydrous citric acid (solution A), 0.4 M anhydrous sodium phosphate dibasic (solution B). For buffer, use 92 ml solution A and 50 ml solution B
- thin coverslip (thickness 0)
- rubber cement or nail varnish
- fluorescence microscope with camera

Method

- 1 Dissolve 0.5 g quinacrine dihydrochloride in 100 ml of distilled water. Use this immediately or store covered in foil in a refrigerator. Make up the MacIlvaine’s buffer (pH 5.6).
- 2 Place the slide in quinacrine stain for 10 min.
- 3 Rinse the slide in tap water.
- 4 Place the slide in buffer for 1–2 min.
- 5 Place a very thin coverslip (thickness 0) over this, and blot excess buffer using filter paper. Seal the edges with rubber cement or nail varnish.
- 6 Analyse using the fluorescence microscope.
- 7 Photograph cells.
If a further banding technique is to be performed on the slides, the coverslips can be removed by cutting through the sealant. The slides can then be rinsed in water and dried.

.....

Troubleshooting

Weakly fluorescing chromosomes

This is caused either by poor illumination or a misaligned microscope, or if too much buffer is present under the coverslip when examination of the slides takes place.

- *Check microscope.*
- *Ensure minimal amount of buffer under coverslip.*

High background fluorescence

This may be caused by the incorrect type of immersion oil being used. Check immersion oil is correct for microscope.

Protocol 23 Reverse banding

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

This method is useful if an abnormality is suspected at the telomeres or in the pale-staining G-band regions. Basically, slides are treated in various buffers at high temperatures. This destroys most of the chromosome structure except the telomeres or pale-staining areas in G-banding. The chemical basis of this is not clear. The method below uses acridine orange as the stain.

Materials

- 0.01% acridine orange
- anhydrous potassium phosphate dibasic (9.93 g l⁻¹) (solution A)
- potassium phosphate monobasic (9.1 g l⁻¹) (solution B)
- water bath at 85 °C
- fluorescence microscope with camera
- Coplin jars

Method

- 1 Prepare 0.01% acridine orange in phosphate buffer (10 mg in 100 ml) Store this in a dark container in the refrigerator.
- 2 Solution A: dissolve 9.93 g l⁻¹ anhydrous sodium phosphate dibasic. Solution B: dissolve 9.1 g l⁻¹ potassium phosphate monobasic.
- 3 Mix 32 ml solution A with 68 ml solution B. Adjust to pH 6.5.
- 4 Add phosphate buffer mixture (A plus B) to a Coplin jar and heat in a water bath to 85 °C.
- 5 Incubate slides for 8–10 min in the heated buffer. (Slides older than 1–2 weeks may require less time.)
- 6 Stain with the acridine orange for 5 min.
- 7 Rinse the slides with buffer.
- 8 Mount the slide with a coverslip (using buffer) and view under a fluorescence microscope using a wavelength of 450–500 nm. Optimum staining shows bands in different gradations between green and red.

9 Photograph if required.

N.B. Optimum R-banding with acridine orange is achieved if slides have aged for 1–2 weeks.

.....

Protocol 24 Constitutive heterochromatin banding

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- 11.05 g l⁻¹ barium hydroxide
- 2×SSC: 17.4 g l⁻¹ sodium citrate, and 8.82 g l⁻¹ sodium chloride. Mix solutions 1 : 1
- Coplin jars
- water bath at 37 °C then 65 °C
- 0.2 M hydrogen chloride (HCl)
- distilled water
- Giemsa: buffer solution (1 : 5)
- DPX mounting medium

Method

- 1 Dissolve 11.04 g l⁻¹ barium hydroxide. Store in an airtight container and filter before use.
- 2 Make up 2×SSC.
- 3 Fill one Coplin jar with filtered barium hydroxide and one with 2×SSC then preheat to 65 °C in a water bath.
- 4 Incubate slides in 0.2 M HCl for 30 min.
- 5 Rinse the slides in distilled water and place in the barium hydroxide for 10 min at 37 °C.
- 6 Incubate the slides in 2×SSC for 2 h at 65 °C.
- 7 Rinse the slides and then stain with Giemsa solution (Giemsa : buffer, 1 : 5) for 20 min.
- 8 Rinse the slides once more, dry and mount in DPX.

Troubleshooting

Barium hydroxide deposits on slide

This is usually caused by precipitation upon exposure to the air. Take the precipitate off the barium hydroxide with a piece of filter paper before inserting slide, and before removal of slide.

Protocol 25 Synchronization technique for bone marrow samples

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- fluorodeoxyuridine: 10 mg in 10 ml distilled water (solution A)
- uridine: 10 mg in 10 ml distilled water (solution B)
- bromodeoxyuridine
- thymidine: 2.5 mg in 10 ml water, use 0.1 ml in 5 ml culture
- complete BM medium (see Table 7.7) or unsupplemented medium
- 12-O-tetradecanoylphorbol-13-acetate (TPA): stock solution of 100 µg ml⁻¹, use a final concentration of 50 ng ml⁻¹ Add 0.1 ml of this to 5 ml culture
- 5 ml reconstituted, lyophilized pokeweed mitogen (PWM); use 0.1 ml per 5 ml culture
- 10 ml reconstituted, lyophilized phytohaemagglutinin (PHA); use 0.1 ml per 5 ml culture

Method

- 1 Set up a 5- to 10-ml culture of bone marrow as described in Section 7.5.1. To this add 0.1 ml of the following cocktail: 30 mg bromodeoxyuridine (BrdU), plus 0.1 ml of fluorodeoxyuridine (FdU) (solution A), 2 ml of uridine (solution B) and make up to 10 ml with distilled water.
- 2 Incubate at 37 °C for 14–17 h.
- 3 Centrifuge at 1000 r.p.m. for 5 min.
- 4 Remove supernatant, and resuspend in complete medium (or unsupplemented medium).
- 5 Repeat step 4 and then add 0.1 ml thymidine (2.5 mg in 10 ml distilled water).
- 6 Incubate at 37 °C for approximately 5 h.

- 7 For the final 15 min of culture add 0.02 µg ml⁻¹ colcemid, and continue processing as in Protocol 13, step 4, but incubate in KCl for 15 min.
-

Acknowledgements

With thanks to Mick Blythe for a smile and ‘Quiet Street’, and Andrei Kanchelskis for his technically excellent feet.

References

1 Rooney, D.E. & Czepulkowski, B.H. (eds) (1992) *Human Cytogenetics: A Practical Approach*. Vol. 1, *Constitutional Analysis* (2nd edn, Oxford University Press, Oxford).

2 Verma, R.S. & Babu, A. (1989) *Human Chromosomes: Manual of Basic Techniques* (Pergamon Press, New York).

3 Brown, W.T., Jenkins, E., Neri, G. *et al.* (1991) Conference report: Fourth International Workshop on the fragile X and X-linked mental retardation. *Am. J. Med. Genet.* **38**, 158–172.

4 Chromosome analysis guidelines—preliminary report. *Report of the Association of Cytogenetic Technologists Task Force* (1990) *Cytogenet. Cell Genet.* **54**, 1–4.

5 ISCN (1991) *Guidelines for Cancer Cytogenetics*. Supplement to *An International System for Human Cytogenetic Nomenclature* (Mitelman, F., ed.) (S. Karger, Basel).

6 Nowell, P.C. & Hungerford, D.A. (1960) A minute chromosome in human chronic granulocytic leukemia. *Science* **132**, 1497–1499.

7 Heisterkamp, N., Stephenson, J.R., Groffen, J. *et al.* (1983) Localization of the c-abl oncogene adjacent to a

translocation breakpoint in chronic myelogenous leukaemia. *Nature* **306**, 239–242.

8 Groffen, J., Stephenson, J.R., Heisterkamp, N. *et al.* (1984) Philadelphia chromosomal breakpoints are clustered within a limited region bcr, on chromosome 22. *Cell* **36**, 93–99.

9 Rooney, D.E. & Czepulkowski, B.H. (1992) Prenatal diagnosis. In *Human Cytogenetics: A Practical Approach*. Vol. 1, *Constitutional Analysis* (2nd edn, Oxford University Press, Oxford), 55–89.

10 Simoni, G., Brambati, B., Danesino, C., *et al.* (1983) Efficient direct chromosome analyses and enzyme determinations from chorionic villi samples in the first trimester of pregnancy. *Hum. Genet.* **63**, 349–357.

11 Flori, E., Nisani, I., Flori, J., Dellenbach, P. & Ruch, J.V. (1985) Direct fetal chromosome studies from chorionic villi. *Prenat. Diag.* **5**, 287–289.

12 Sumner, A.T. (1986) Mechanisms of quinacrine binding and fluorescence in nuclei and chromosomes. *Histochemistry* **84**, 566–574.

13 Benn, P.A. & Perle, M.A. (1991) Chromosome staining and banding techniques. In *Human Cytogenetics: A Practical Approach*. Vol. I, *Constitutional Analysis* (2nd edn, Oxford University Press, Oxford), 91–118.

14 Czepulkowski, B.H. (1994) Malignancy and acquired chromosome abnormalities. In *Human Cytogenetics: Essential Data* (Czepulkowski, B.H. & Rooney, D.E. eds) (BIOS/John Wiley, Oxford).

15 Mitelman, F. (1994) *Catalog of Chromosome Aberrations in Cancer* (5th edn, Wiley-Liss, New York).

Chapter 8

Solid tumour cytogenetics

Sandra Birdsall, Yong-Jie Lu & Janet Shipley

Molecular Cytogenetics Laboratory, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK

8.1	Introduction, 189		
8.2	Techniques for chromosome preparation, 190		
8.2.1	Tumour collection, 190		
8.2.2	Disaggregation and washing, 190		
8.2.2.1	Mechanical disaggregation, 190		
8.2.2.2	Enzymatic disaggregation, 191		
8.2.3	Direct preparations, 191		
8.2.4	Establishing short-term cultures, 191		
8.2.5	Establishing long-term cultures, 191		
8.2.6	Maintenance of short- and long-term cultures, 193		
8.2.7	Harvesting, 193		
8.3	Alternative strategies using fluorescence in situ hybridization, 193		
8.3.1	Comparative genomic hybridization, 193		
8.3.1.1	Analysis of CGH, 194		
8.3.2	Interphase FISH, 195		
8.4	Final comments, 195		
	Protocol 26		
	Mechanical disaggregation, 196		
	Protocol 27		
	Enzymatic disaggregation, 197		
	Protocol 28		
	Direct chromosome preparations, 198		
	Troubleshooting, 199		
	Low number and poor morphology of metaphase spreads, 199		
	Protocol 29		
	Establishing short-term cultures from cell suspensions, 199		
	Troubleshooting, 200		
	Optimum cell density, 200		
	Protocol 30		
	Establishing short-term cultures from explants, 201		
	Troubleshooting, 202		
	Establishing cell cultures from explants, 202		
	Protocol 31		
	Lipofectin-mediated transfection for establishing long-term cultures, 202		
	Troubleshooting, 203		
	Establishing a cell line, 203		
	Protocol 32		
	Passaging cells, 203		
	Protocol 33		
	Harvesting by removal of adherent cells, 204		
	Protocol 34		
	Harvesting in situ, 205		

Troubleshooting, 206

Unsatisfactory metaphases, 206

Protocol 35

Labelling genomic DNA by nick translation for CGH
and hybridization for CGH, 206

Troubleshooting, 207

Standardizing conditions, 207
Weak or grainy signals, 208

Protocol 36

Tumour imprint preparation, 209

Protocol 37

Pretreatment of tumour imprints prior to
hybridization, 209

Troubleshooting, 210

Ensuring slide quality, 210

Acknowledgements, 210

References, 211

8.1 Introduction

The concept of the genetic basis for the initiation and progression of probably all tumours is now widely accepted. Consistent, specific chromosome rearrangements that are associated with particular tumour types are a visible hallmark of the genetic changes and as such have been instrumental in defining which regions of the genome are altered in some tumour types. In some cases, this has led to identifying which genes, at a specific genomic location, are implicated in the pathogenesis of the tumours (see Appendix IX), which in turn has led to determination of the molecular mechanisms involved.

Chromosome rearrangements include deletions, translocations, and gain or amplification of chromosomal segments. Consistent loss of a region may indicate the site of a tumour suppressor gene. In a translocation, one part of a chromosome becomes joined to another. At the molecular level, a gene located at or near the breakpoint on one of the chromosomes is fused to sequences from the other chromosome. This most commonly results in the formation of a fusion gene and production of a fusion protein but can also place a gene under the control of a novel regulatory element [1]. Amplification of a region is apparent cytogenetically as structures known as double minutes (dms) and homogeneously staining regions (hsrs) (see Chapter 7, Table 7.5 for abbreviations used in cytogenetics). The over-representation of genes associated with these structures results in their overexpression. All these events are believed to play an important role in the development and progression of the tumours.

Rearrangements and their associated molecular events may be unique markers for a particular tumour type and may therefore be a useful diagnostic aid in cases where the differential diagnosis of a particular tumour is important. It has already been suggested for some groups of tumours that the presence of particular rearrangements may be a more rational means for classification than present immunohistological criteria. The molecular events may also provide novel targets for new therapeutic strategies. In addition, chromosome rearrangements may also be associated with different prognostic groups and may therefore be of clinical significance.

Chromosome rearrangements in solid tumours can be divided into three categories, some or all of which may be apparent in a given sample:

1 Primary changes Simple clonal abnormalities, sometimes found as the sole rearrangement and which are associated with specific tumour types or subtypes.

2 Secondary changes Rearrangements that are less specific than primary changes and occur in addition to primary changes. They can be thought to give cells a proliferative advantage and play a part in tumour progression.

3 Cytogenetic noise Non-clonal complex abnormalities which may reflect an unstable genome.

Many of the karyotypes found in solid tumours are extremely complex, particularly for the more malignant tumours and in common cancers such as carcinomas of the breast and colon. In such cases it can be difficult to determine into which of the above three categories the rearrangements fall. The complexity of some of the abnormalities can also make them difficult to characterize accurately. However, rearrangements found in the soft tissue sarcomas in particular are considered to be primary rearrangements and great progress has been made recently in their characterization. They can now be used diagnostically in a manner analogous to the chromosomal abnormalities long known in the haematological malignancies (see Chapter 7, and Appendix IX, Tables IX.1–IX.12). In the region of 6000 karyotypes of solid tumours have been reported, which represent approximately a quarter of all the karyotypes reported for malignancies [2]. Karyotype analysis of solid tumours has lagged behind that of the haematological malignancies mainly for technical reasons, although the proportion of solid tumour cases has been steadily increasing over the last 10 years or so. Improved methods of tissue disaggregation, culture of the tumour cells and chromosome spreading have all played their part in this increase. The use of chromosome banding described in Chapter 7 plays an integral part of the preparation of the chromosomes for analysis. However, generally speaking, the preparation of chromosomes from solid tumours is more difficult and time consuming than in the haematological malignancies.

This chapter details some of the methods commonly adopted for the preparation of chromosomes from tumours derived from different tissues. In addition, important new approaches, which negate the need for chromosome preparation, have emerged over the past few years. Two techniques based on fluorescence *in situ* hybridization (FISH) (see Chapter 9) will be described here. Comparative genomic hybridization (CGH) involves identifying gains (including genomic amplification) and losses of chromosomal material following the cohybridization of differentially labelled tumour and normal DNA to normal chromosomes [3,4]; interphase FISH analysis utilizes region-specific markers to identify specific rearrangements in non-dividing cells [5,6].

The sorting of chromosomes by flow cytometry can also be used to distinguish aberrant chromosomes (see Chapter 12).

8.2 Techniques for chromosome preparation

There are many different variations on the techniques that have been successfully used to prepare chromosomes from tumours, some of which are described below. The success rates for different tumour types varies enormously: for example, the overall success rate for culturing neuroblastomas is poor and in the region of 30%. Also, as individuals gain experience and a feel for the methods, the probability of producing successful chromosome preparations and the quality of these usually improves.

The first practical consideration is collection and transportation of the specimens from source to the laboratory for processing. The methods for chromosome preparation usually involve the disaggregation of tumour cells into a suspension which can be done either mechanically or, more usually, by enzymatic treatment. The most commonly used approach is that of short-term culture (several days to weeks) although direct preparations and long-term cultures are also possible. In some cases it may be more appropriate to set up explant cultures which would not require the production of cell suspensions. Harvesting dividing cells can either involve the removal of adherent cells from tissue culture vessels prior to hypotonic treatment and fixation or harvesting *in situ*. The latter is particularly appropriate for use with small numbers of cells.

8.2.1 Tumour collection

The condition of the cells within the tumour sample is an important factor in the ability to produce good-quality chromosome preparations. It is advisable to remove any necrotic areas and to keep the sample as sterile as possible. Tumour samples should be

placed in sterile containers with an air-buffered medium or other tissue culture media—for example, Leibovitz's L15 (L15) (air-buffered), or RPMI 1640 with 25 mM Hepes, supplemented with 100 units ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin. If these are not available, phosphate-buffered saline (PBS) or isotonic saline will suffice. Samples can be kept in L15 or medium for several days without unduly compromising the potential to culture the cells. Samples that may be contaminated with bacteria or fungi—for example, those from head and neck tumours—should be exposed to additional antibiotics (e.g. gentamycin 0.5–50 µg ml⁻¹, neomycin sulphate 50 µg ml⁻¹) and antifungal agents (e.g. Fungizone (amphotericin B) 0.25–25 µg ml⁻¹), both in the transporting medium and in the disaggregation steps.

Fine-needle aspirates should be collected into medium containing 10 units per litre of heparin. The temperature of the medium into which the tumour is placed can be important for successful culture, for example, fat-derived tumours should be kept as close as possible to 37 °C prior to processing. Large specimens may be transported dry in, for example, a plastic bag. It may be advisable to store part of the sample at –80 °C or in liquid nitrogen for future molecular studies. For RNA isolation, material snap frozen as soon after removal as possible is best. In addition, it is worth considering whether a sample of normal tissue or blood is required. Before processing the sample, any normal tissue, particularly fat, or any remaining necrotic regions should be trimmed away.

8.2.2 Disaggregation and washing

Although the advent of enzymatic disaggregation, particularly the use of collagenase, has led to the large increase in successful karyotyping of solid tumours, it is not always necessary, and mechanical disaggregation alone may be sufficient for some tumour types. In fact, for some tumour types, mechanical disaggregation can produce a greater number of successful cultures, for example in small round-cell tumours of childhood [7]. There is also some evidence that the method of disaggregation affects the type of cells released and hence the karyotype obtained [8].

8.2.2.1 Mechanical disaggregation

Single cells and small clumps released by mechanical disaggregation can be used either to initiate short-term cultures or for direct harvesting. If the sample is being used for direct harvesting alone colcemid can be added at a final concentration of

Solid tumour cytogenetics is used to:

- define consistent rearrangements associated with particular tumour types and subtypes

This can be used to:

- identify genes associated with tumour initiation and progression
- aid diagnosis and indicate prognosis

0.1µgml⁻¹ to culture media during processing. Protocol 26 describes a method for mechanical disaggregation.

8.2.2.2 Enzymatic disaggregation

The large increase in successful karyotyping of solid tumours has been brought about in the main by the use of enzymes to disaggregate the tissue. The most commonly used enzymes are the collagenases, which cleave the peptide bonds in intercellular collagen. Crude collagenases are commercially available which contain a mixture of collagenase, non-specific protease and clostripain. Several different types of collagenases with different molecular weights have been isolated which preferentially release different cell types. The type, concentration and incubation time used for successful collagenase digestion varies greatly between tumour types and is indicated in Table 8.1. Sometimes, particularly for tumours with mixed cell populations, a variety of collagenase types should be tried [9]. It is advisable to minimize enzymatic treatment of the tissue due to the potential damage it may cause the living cells; therefore after collagenase incubation the tumour pieces are further dissociated by passing several times through a Pasteur pipette. DNase I (0.02%), hyaluronidase (0.01%), and pronase (0.05%) can be added to the collagenase solution to digest any remaining intercellular material. This is particularly useful for fibrous tissues. Protocol 27 describes a method for enzymatic disaggregation using collagenase.

8.2.3 Direct preparations

This approach is only appropriate for tumours with a high rate of mitotic activity and therefore is likely

to be most successful with rapidly growing malignant tumours. The quality of the preparations is usually inferior to those prepared by short-term culture but is a true representation of the population of the dividing cells *in vivo*. An example of a karyotype from a direct preparation of an aggressive breast carcinoma is shown in Fig. 8.1. Protocol 28 describes a method for direct chromosome preparations.

8.2.4 Establishing short-term cultures

The use of short-term culture has largely developed from the improved methods for the disaggregation of cells into a single-cell suspension and the use of specialized tissue culture media, in some cases supplemented with growth factors and other additives. Various sizes of commercially available flasks and dishes are available enabling culture of a range of different cell numbers. By using chambers adhered to microscope slides and coverslips in multiwell plates, combined with *in situ* harvesting techniques (see Protocol 34), it is possible to produce chromosome preparations from minimal amounts of starting material. In addition to using disaggregated cells, it is sometimes appropriate to grow tumour cells from explants—for example, squamous cell carcinomas. Two examples of cells in short-term culture are shown in Figs 8.2 and 8.3. Protocol 29 describes a method for establishing short-term cultures from cell suspensions and Protocol 30 a method for establishing short-term cultures from explants.

8.2.5 Establishing long-term cultures

After a relatively small number of divisions, many

Table 8.1 Collagenase concentration and exposure times for different tumour types.

Tumour type	Concentration (U ml ⁻¹)	Time (h)	Collagenase type
Soft tissue tumours	1000	1–2	II/IV
Uterine leiomyomas	1000–2000	15–24	II/IV
Lipomas	100–200	15–24	II/IV
Breast sarcomas	1000	2–4	I/II/IV
Breast carcinomas	900	24	I
Renal cell carcinomas	1000	0.5–1	II
Bone tumours	1300–1500	2–4	II
Brain tumours	1300–1500	2–4	II
Gastrointestinal tumours	900–1500	1–5	II
Lung tumours	200–400	15–24	II
Prostatic tumours	200–900	15–24	I/IV
Ovarian tumours	200–1500	2–4	II/III
Germ cell tumours of testis	1000–1500	1–24	II
Head and neck tumours	200–400	15–24	II

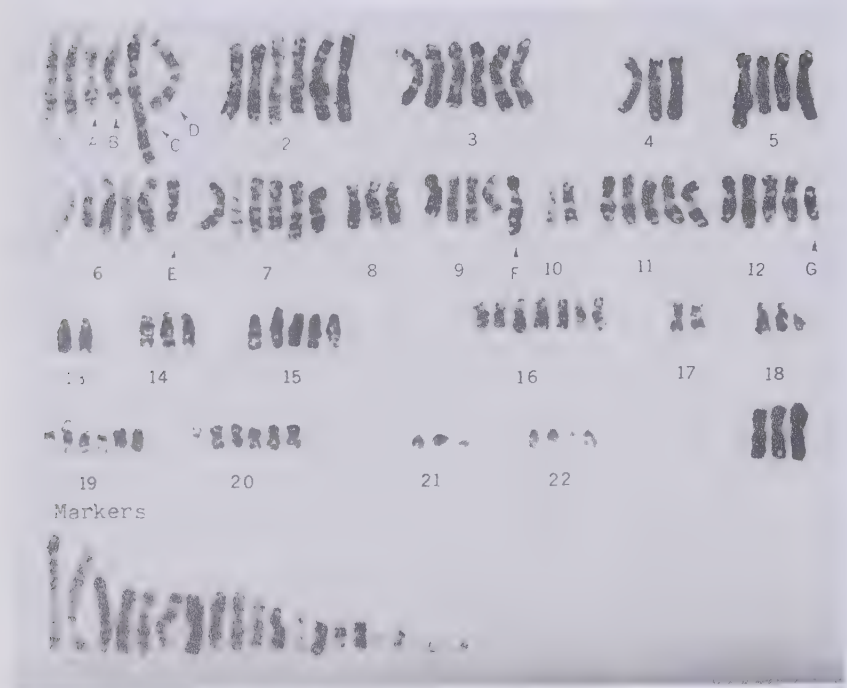


Fig. 8.1 G-banded karyotype of a directly harvested breast carcinoma.



Fig. 8.2 Photomicrograph of cells in short-term culture from a phyllodes tumour of breast.

cell types stop dividing in culture. Establishing long-term (immortal) cultures is often desirable for functional and other studies. Although cell lines are a good source of metaphase chromosomes, they often have a more complex karyotype than the original sample. Generally, the original chromosome abnormalities are retained within a background of culture-induced rearrangements. A small proportion of short-term cultures of tumour cells, if main-

tained for a sufficient length of time, will continue to grow and can be considered to be immortal.

Viral genes—for example, the early region of SV40—are usually used to transform cells. Although cells can be transfected directly by viral infection, most people transfect the cells with plasmids containing the early region of SV40. Such plasmids also carry a selectable marker such as the gene for neomycin resistance, and can thus be selected by G418 sulphate. The plasmid is introduced into the cells using a variety of methods such as calcium phosphate-mediated transfection, electroporation, protoplast fusion, and liposome-mediated transfection systems [10]. Lipofectin is a commercially available liposome reagent suitable for transfecting nucleic acids into tissue culture cells and Protocol 31 describes its use. Some time after transfection the culture goes through a crisis period when most cells die. The remaining cells eventually start to divide more rapidly and can be considered immortal.

Some researchers transplant tumour cells into nude mice as a way of maintaining the cells. Although the tumour cells grow well in nude mice, the cells obtained are just as difficult to produce chromosomes from as the original sample. Such samples have to be processed for chromosome analysis as in Protocols 26 and 27. Care should be taken to make sure that the cells are not contaminated by mouse cells. Once a cell line is established this way, confirmatory karyotype analysis is very important.

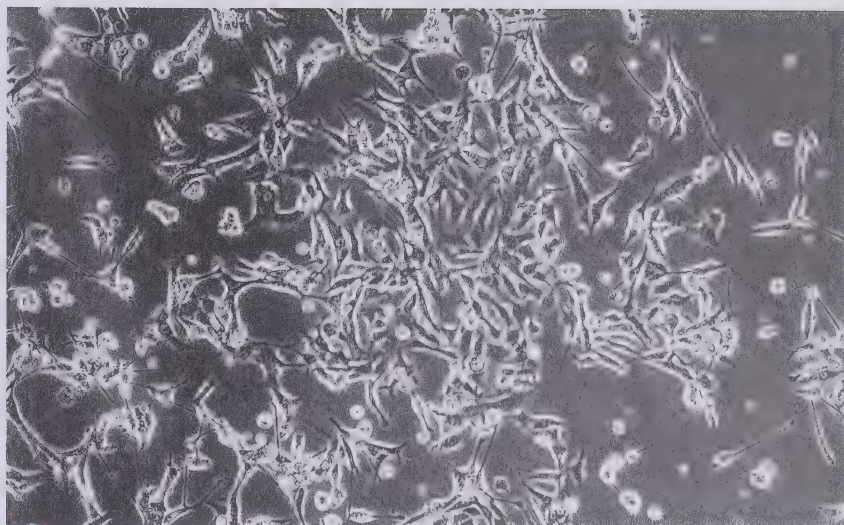


Fig. 8.3 Photomicrograph of cells in short-term culture from a renal cell adenocarcinoma.

8.2.6 Maintenance of short- and long-term cultures

Most actively growing cells require a change of medium once or twice a week. Initially, however, it is advisable to disturb the cultures as little as possible beyond initial removal of unattached cells. If the cells become near-confluent they can be passaged as described in Protocol 32.

8.2.7 Harvesting

Cultures should be observed daily to determine the optimum time to harvest dividing cells. Several mitoses per low-power field is a good indication that the cultures are ready; colcemid is then added directly to the culture and it is incubated further. Colcemid inhibits spindle formation and blocks cells in mitosis. It can be added for either a shorter time at a higher concentration or a longer period at a lower concentration. For short-term colcemid exposure, 1–2 h at a final concentration of $0.1 \mu\text{g ml}^{-1}$ colcemid, and for long-term exposure 16–18 h at a final concentration of $0.01 \mu\text{g ml}^{-1}$ colcemid is appropriate.

The best chromosome harvests are usually accomplished in the first few days of culture when tumour cells are likely to be progressing through the cell cycle most rapidly and fibroblast contamination is at a minimum [11]. For some tumour types, particularly epithelial tumours, it is often possible to remove normal fibroblasts using differential trypsinization. As epithelial cells require a longer trypsinization time than fibroblasts, the cells can be incubated in trypsin/EDTA for 2–3 min, the loose cells removed and discarded and fresh medium added.

Cells can be harvested either by removal of adherent cells and by processing in a fashion analogous to harvesting blood cultures (see Chapter 7) or harvested *in situ* when the cells have been grown in slide flasks or on coverslips. Protocol 33 describes a method for harvesting by removal of adherent cells and Protocol 34 describes a method for harvesting cells *in situ*.

Occasionally, for rapidly dividing samples with a large number of metaphases, it is not necessary to remove all the cells at harvesting and a mitotic shake is used. After exposure to colcemid the flask is tapped sharply to release any dividing cells. The medium is removed to a centrifuge tube and processed as in Protocol 33. The remaining cells in the flask can then continue to grow as fresh medium is added. An example of a partial karyotype prepared from harvesting a short-term culture is shown in Fig. 8.4.

8.3 Alternative strategies using fluorescence in situ hybridization

In addition to using chromosome analysis to obtain karyotypic information, important new approaches which negate the need for chromosome preparation have emerged over the last few years. Two of those based on FISH (see Chapter 9) will be described here. CGH identifies gains and losses of chromosomal material [3,4] and interphase FISH analysis can be used to identify specific rearrangements in nondividing cells [6].

8.3.1 Comparative genomic hybridization

CGH is a method in which DNA from a tumour and

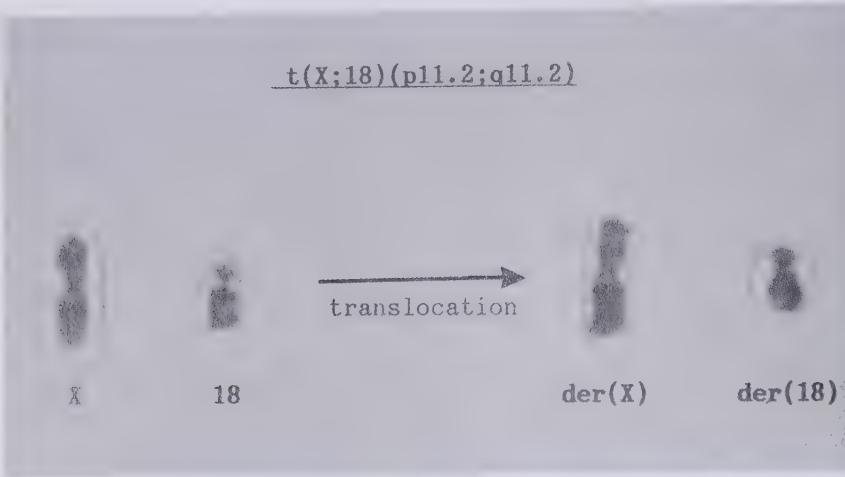


Fig. 8.4 Partial karyotype of a $t(X;18)(p11.2;q11.2)$ associated with synovial sarcoma.

a normal control are differentially labelled and cohybridized to normal metaphases. An abnormal ratio between the intensity of signals from each DNA sample is indicative of a regional copy number change [3,4]. It is also possible to hybridize only the tumour DNA and to use abnormally bright areas as an indication of an amplified region without the use of reference DNA [12]. DNA for CGH can be prepared from microdissected regions of a tumour including paraffin-embedded material. In conjunction with this it may be necessary to use a polymerase chain reaction (PCR)-based approach with degenerate primers (DOP-PCR; see Chapter 11) in order to generate sufficient quantities of DNA for CGH experiments [13]. CGH analysis of the gains and losses of material in a sample can provide complementary data to karyotype information [14]. It can also lead to the identification of which genes are amplified in a tumour sample (see Case Study 8.1) [15].

The preparation of normal chromosomes, either with or without incorporation of bromodeoxyuridine to enhance banding, is described in Chapters 7, 9 and 11. The CGH method involves preparing DNA samples by standard methods [10] and labelling these, preferably directly with fluorescently labelled nucleotides, in the manner described in Chapter 9, with the modifications described in Protocol 35. This is followed by a long period of cohybridization of the labelled DNAs to denatured chromosomes including Cot-1 DNA to suppress the hybridization of repetitive elements, and finally slide washing to remove material which has not hybridized. The protocol is based on the method described by Kallioniemi *et al.* [16].

8.3.1.1 Analysis of CGH

The signals can be visualized using a good fluores-

Novel formation and amplification of the *PAX7-FKHR* fusion gene in a case of alveolar rhabdomyosarcoma [15]

Alveolar rhabdomyosarcoma frequently exhibits double minutes which are evidence of genomic amplification (see Fig. 8.5) and have specific translocations that result in the fusion of the *FKHR* gene at 13q14 with either the *PAX3* gene at 2q35 or, more rarely, the *PAX7* gene at 1p36. Comparative genomic hybridization revealed amplification at 13q14 and 1p36, suggesting amplification of the *PAX7-FKHR* fusion gene in two cases of alveolar rhabdomyosarcoma (see Plate 1). A *PAX7-FKHR* fusion transcript was demonstrated in both cases by reverse transcription PCR followed by sequence analysis. In one case, amplification of the *PAX7* gene and 3' and 5' *FKHR* gene sequences was demonstrated using interphase fluorescence *in situ* hybridization (FISH) on tumour imprints (see Plate 2). The colocalization, variable copy number and distribution of signals in nuclei was consistent with amplification of these sequences on double minutes, which were present cytogenetically. Chromatin release studies (see Section 9.2.3) suggested that the amplified *PAX7-FKHR* fusion gene resulted from the insertion of *PAX7* sequences into the first intron of the *FKHR* gene, in keeping with the absence of cytogenetic evidence for derivative chromosomes.

Case Study 8.1

cence microscope with appropriate filters. Obvious differences in intensity at a particular locus, including those seen in single colour experiments, may be apparent. Capture of images in a digital format is increasingly carried out using charged-coupled devices (CCDs) and cooled CCDs attached to computer systems. This allows image processing and an excellent facility for record keeping. Registration of the images from different filters is an important consideration and potential problems can be overcome using novel beam splitters and

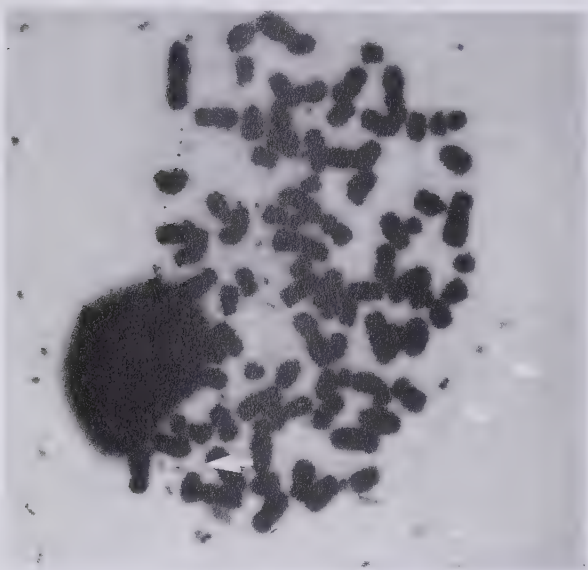


Fig. 8.5 Chromosomes from an alveolar rhabdomyosarcoma block stained to show double minutes (dmin) as indicated by arrows (see also Plates 1 and 2).

emission filters and placing the excitation filters adjacent to the light source (see Chapter 13). Many systems are commercially available for the digital capture and analysis of fluorescent signals. Most of these incorporate the necessary algorithms to compare the fluorescence ratios along the length of chromosomes and offer ways of scoring and displaying the data. Different approaches to measuring the fluorescence ratios are possible but will not be discussed here [3,4]. Digital imaging is discussed further in Chapter 13. An example is shown in Plate 1. It is important to establish the normal variation in the fluorescence ratios in order to determine which regions have an abnormal ratio. This can be done by either comparing the signal ratios following hybridization of differentially labelled normal DNA to chromosomes or by using a chromosome region in the test DNA which is known, by some other means, to have a normal copy number.

8.3.2 Interphase FISH

Occasionally it is not possible to obtain metaphases from solid tumours due to growth failure, the small size, or the inappropriateness of the sample received (i.e. snap-frozen material or paraffin-embedded tissue). Alternatively, one may be only looking for a very specific rearrangement or copy number change, possibly in a short time frame. In such cases FISH can be performed on interphase nuclei (see Chapter 9, Section 9.2.2).

Nuclei can be obtained from a cytogenetic harvest that failed to yield any metaphase chromosomes, from nuclei fixed immediately after collagenase treatment, from tissue sections cut from paraffin-embedded tissue [17], from nuclei released from paraffin sections [18] or from tumour imprints made from fresh or frozen samples. A variety of probe types and differentially labelled mixtures of these can be hybridized to the nuclei; for example, chromosome-specific centromere and painting probes [19], and region-specific markers flanking a translocation breakpoint [6,20]. Protocol 36 describes a method for preparing tumour touch imprints; the conditions required for hybridization to these and other slide preparations derived from fresh material are indicated in Protocol 37.

8.4 Final comments

Solid tumours are highly variable both within and between types, and therefore there are many different conditions and protocols which are appropriate for successful karyotyping. Some tumour types disaggregate easily by purely mechanical methods, others require extended collagenase incubations at high enzyme concentrations and some may grow best as explants. Although most tumours will grow in commercially available media, there is increasing evidence that reducing the amount of fetal calf serum used and the addition of a variety of specific growth factors increases the success rate. The culture method and conditions can not only affect the chances of obtaining a tumour karyotype but can also influence the actual chromosome abnormalities seen. Some solid tumours, particularly epithelial tumours, appear to contain several unrelated abnormal clones which *in vitro* can become artificially dominated by one clone and thus bias the results [21]. The new FISH-based approaches using interphase nuclei and CGH, potentially in conjunction with microdissection (see Chapter 11), avoid the problems of *in vitro* growth selection.

It is often difficult to obtain good-quality metaphase spreads from solid tumours and the karyotype of many solid tumours can be very complex. However, the value of the information obtained from a full or partial karyotype analysis should not be underestimated. Rearrangements can be confirmed by hybridizing region-specific markers, chromosome-specific centromere and painting probes to the tumour's chromosomes. CGH and interphase can be used to give complementary information to traditional karyotype data or used in situations where it may not be possible to produce a karyotype. CGH analysis does not indicate rear-

rangements that do not alter the copy number such as translocations and inversions, and is dependent on the DNA isolated which may be non-clonal in origin. Interphase FISH can only be used to look for specific rearrangements or copy number changes in samples.

These approaches for gaining karyotype infor-

mation are complementary and have their place in providing cytogenetic information for a given tumour. Together, they should continue to make an important contribution towards the identification of genes implicated in the pathogenetic process and an increasingly important role in tumour diagnosis and the management of patients.

Table 8.2 Commonly used media and supplements.

Tissue type	Media	FCS concentration	Supplements
Mesenchymal tumours	F12	10–20%	
Synovial sarcoma			Epidermal growth factor 2.5 ng ml ⁻¹
Ewing's sarcoma	F12	10–20%	Insulin 4 µg ml ⁻¹
Breast sarcoma	DMEM/F12	10%	
Breast carcinoma	RPMI	10%	Cholera toxin, insulin, hydrocortisone (excellent for normal breast)
	CDM-5 [23]		
	DFCI1 [24]		
Head and neck tumours	RPMI	10%	Cholera toxin 0.1 µg ml ⁻¹ , insulin 4 µg ml ⁻¹ , epidermal growth factor 2.5 ng ml ⁻¹ , Hepes 10 mM, Fungizone (amphotericin B) 2.5 µg ml ⁻¹
Renal tumours	RPMI 1640	17%	Hydrocortisone 0.36 µg ml ⁻¹
Primitive neuroectodermal tumours of the CNS	DMEM	20%	
Lung tumours	MCDB151 [25]		
Germ cell tumours	RPMI	10%	

Protocol 26 Mechanical disaggregation

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.

Materials

- medium: L15 supplemented with penicillin and streptomycin at 100 units ml⁻¹ and 100 µg ml⁻¹, respectively. See also Table 8.2 for commonly used media for different tumour types

Method

- 1 Wash the specimen in 2–3 changes of medium (L15 supplemented with penicillin and streptomycin at 100 units ml⁻¹ and 100 µg ml⁻¹, respectively). Transfer to a sterile Petri dish and remove any obviously normal or necrotic tissue using scalpels.
- 2 Remove a representative piece of tumour for snap freezing if required. Tumour touch imprints can also be made from this piece before freezing (see Protocol 36).
- 3 Large specimens should be divided into approximately 1-cm³ pieces, kept moist in L15 and processed separately.
- 4 Add a small amount of culture medium (1–2 ml) (see Table 8.2) to keep sample moist while mincing. Mince the tissue finely using two scalpels until tissue has fully disaggregated or fragments of 1–2 mm³ are achieved.
- 5 Transfer the minced tissue and medium to a universal container using a wide-bore transfer pipette.
- 6 Leave to sediment for about 5 min. Remove the supernatant and place in a centrifuge tube. This will contain any single cells or small clumps released by the purely mechanical disaggregation. The remaining lumps can then be further disaggregated by collagenase digestion (see Protocol 27).
- 7 Spin the supernatant at 300 g for 5 min.
- 8 Remove the medium, resuspend cell pellet in fresh culture medium and distribute to tissue culture flasks or centrifuge tubes as appropriate.

Protocol 27 Enzymatic disaggregation

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.

Materials

- collagenase (see Table 8.1 for details)

Method

- 1 Follow Protocol 26 for the initial mechanical disaggregation of the tumour.
- 2 Resuspend the minced sediment in culture medium (see Table 8.2) containing the appropriate concentration of collagenase (see Table 8.1). Transfer to a universal container.

- 3 Incubate the fragments in the collagenase medium for the appropriate length of time at 37 °C (see Table 8.1). During this time agitate the contents of the universal several times by swirling the mixture. The required end-point for cell culture is a mixture containing small clumps of cells rather than single cells. However, for direct harvests a single-cell suspension is preferable.
- 4 At the end of the collagenase incubation period further dissociate the tissue fragments by pipetting. It may be advisable to exclude larger undigested fragments by settling.
- 5 Spin in a bench centrifuge at 300 g for 5 min.
- 6 Discard supernatant, add fresh culture medium, mix well and spin.
- 7 Repeat twice.
- 8 Discard supernatant, resuspend in appropriate amount of culture medium and distribute into centrifuge tubes for direct harvesting or tissue culture flasks for long- or short-term culture.

.....

Protocol 28 Direct chromosome preparations

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.
(See also Chapter 7, Protocol 13 and Chapter 11, Protocols 57 and 58.)

Materials

- colcemid stock solution (10 µg ml⁻¹)
- 0.075 M KCl
- fixative (methanol/glacial acetic acid, 3 : 1)
- centrifuge tubes
- clean glass slides

Method

- 1 After mechanical disaggregation alone or combined with enzymatic disaggregation, described in Protocols 26 and 27, resuspend all or part of the pellet in culture medium in a centrifuge tube. If any visible lumps remain, allow to settle and use only the resuspended cells for direct harvesting.
- 2 Add colcemid to a final concentration of 0.1 µg ml⁻¹.
- 3 Incubate for 2–4 h at 37 °C.
- 4 Centrifuge at 300 g for 5 min.
- 5 Add 5–10 ml prewarmed 0.075 M KCl.

- 6 Incubate 5–10 min at 37 °C.
- 7 Centrifuge at 300 g for 5 min.
- 8 Remove supernatant and resuspend pellet by flicking.
- 9 Add 5–10 ml fixative carefully dropwise, mixing constantly.
- 10 Change fixative 1–3 times to remove cell debris. (This can be checked by making slides. For methods of slide making, refer to Protocol 13 in Chapter 7).

Troubleshooting

Low number and poor morphology of metaphase spreads

Often the low number and poor morphology of metaphase spreads obtained in direct preparations is a feature of the tumour itself and short-term cultures may help.

Improvements may be obtained by altering the exposure time to hypotonic solution, as tumour types vary considerably in the optimum KCl incubation time required. For soft tissue tumours the average time is 8 min; however, if a low mitotic index is found in the first harvest, the KCl incubation time may be profitably reduced for subsequent harvests. For many solid tumours one change of fix is adequate to produce clean preparations that spread well, but sometimes further fixation may be required to clear cell debris. This can be monitored by dropping the fixed cell suspension onto a clean microscope slide after each fixation. Generally, reducing the number of fixation steps minimizes the loss of dividing cells during preparation. An alternative method for fixation is to add 1 ml of fresh fixative to the tube after KCl incubation before centrifugation at step 4. This can reduce the clumping of cells which may trap the metaphases and aids the removal of cytoplasm.

Protocol 29 Establishing short-term cultures from cell suspensions

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.

Materials

- 25-cm² flasks or chamber slide flasks (e.g. Nunc flaskettes) or coverslips

Method

- 1 Obtain a cell suspension using either mechanical or enzymatic disaggregation as detailed in Protocols 26 and 27.
- 2 Since it is preferable to use small clumps of cells rather than single-cell suspensions to inoculate flasks, traditional methods of determining viable cell numbers for plating purposes are invalid. The size and number of cells present in a tumour can vary enormously. However, as a rough guide, one or two 25-cm² flasks can be set up from 1 cm² of tissue. For smaller cell numbers, chamber slide flasks (e.g. Nunc flaskettes) or coverslips are better for establishing a culture.
- 3 Cells that grow as monolayers often adhere best when placed in the flask in a small volume of medium. For a 25-cm² flask, 2–2.5 ml of medium rather than the usual 5 ml is advisable. Similarly, for a chamber slide flask or small coverslip in a dish, a drop (<1 ml) of medium rather than the usual 2.5 ml is best in the initial phase.
- 4 Gas and cap tightly, or place flasks with loose caps or dishes in a gassing incubator.
- 5 Incubate at 37 °C.
- 6 Observe cultures the next day and if necessary change the medium. Large numbers of reactive lymphocytes are often present in the original cell suspension; these rapidly exhaust the medium and should be removed as soon as the tumour cells have settled.

Culture media and additives

The choice of tissue culture media is related to tumour type (see Table 8.2), with the general principle that RPMI 1640 is more appropriate for epithelial cells and DMEM is good for cells of mesenchymal origin. Such standard media are generally supplemented with L-glutamine (2 mM), penicillin (100 U ml⁻¹), streptomycin (100 µg ml⁻¹) and fetal calf serum (10–20%). For individual tumour types, further supplements can be added which reduce the amount of serum required. Low-serum or serum-free media have the advantage of reducing the overgrowth of normal fibroblasts; they can, however, also reduce the proliferation of the tumour cells.

Troubleshooting

Optimum cell density

The density of viable cells seeded into a culture vessel is critical to establishing a viable culture. Too sparse and the cells will not grow, but too dense and the medium will be rapidly exhausted, inhibiting cell

adhesion. Establishing optimum cell density is largely a matter of experience and varies enormously between different tumour cell types and the condition of the tumour sample. If possible, set up a range of different densities. Fat cells may inhibit the growth of the culture and therefore it is important to remove tissue containing such contaminating cells before disaggregation. A significant problem with most short-term cultures is the potential overgrowth of fibroblasts from the stroma. The more confluent the tumour cells in the initial culture, the less chance the fibroblasts have to establish. Modification to the media used and monitoring to optimize the time of harvesting should minimize contamination by these cells. Differential trypsinization can be used to enrich for cells free from fibroblast contamination (see Section 8.2.7).

.....

Protocol 30 Establishing short-term cultures from explants

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.

Materials

- culture medium (see Table 8.2)
- tissue culture flasks

Method

- 1** The sample is washed, and cut into pieces of 1 mm².
- 2** These are placed in the empty tissue culture vessel and left for a short while to adhere. If it is possible to orientate the explants, the outer epithelium should be uppermost. Scoring the base of the tissue culture flask can aid attachment.
- 3** Before the explants dry, a very small amount (one drop) of medium should be placed on them. The flasks are then gassed or left in a CO₂ incubator.
- 4** The flasks should be left undisturbed for several days until cells are seen to be growing out from the explant. Once there is a reasonable area of cells surrounding the explant, more medium can be added.
- 5** It takes in the region of 2 weeks for the flask to get sufficiently full to passage the cells. Following removal of the cells, medium can be returned to the original flask and further cells allowed to grow out from the explant.

.....

Troubleshooting

Establishing cell cultures from explants

The main difficulty is getting the conditions right to adhere the explant to the tissue culture vessel. If medium is added too soon the explants will float off, and if added too late the cells will have dried up. In either case there will be no cell growth. If the size of the tissue piece is too great then the cells will fail to get enough nutrition and die. The tissue should be cut cleanly to avoid damage to the surface of the explant. Not all the explants will necessarily be in the correct orientation to allow cell growth, but by chance or skill there should be sufficient to establish a culture.

Protocol 31 Lipofectin-mediated transfection for establishing long-term cultures

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.

Materials

- Lipofectin reagent (Gibco-BRL)
- PBS
- G418 sulphate (Geneticin, Gibco-BRL)
- serum-free medium as appropriate (see Table 8.2)

Method

- 1 Dilute 5 µg of the plasmid containing the SV40 early region in 100 µl serum-free medium.
- 2 Dilute the Lipofectin reagent (Gibco-BRL) in 100 µl serum-free medium at a variety of concentrations (such as 1 : 1, 1 : 2, 1 : 4, 1 : 8, wt/wt).
- 3 Mix the diluted plasmid and Lipofectin reagent.
- 4 Incubate at room temperature for 15 min.
- 5 Remove the growth medium from the cells to be transformed and wash twice with PBS.
- 6 Add the plasmid–Lipofectin reagent mixture to the cells.
- 7 Add 800 µl serum-free medium to the culture immediately. Incubate overnight at 37 °C.
- 8 Add 1 ml culture medium the following day.

9 The transfected cells can be selected by the addition of G418 sulphate (Geneticin, Gibco-BRL) to the culture medium which is changed regularly.

.....

Troubleshooting

Establishing a cell line

Establishing a tumour cell line can take a long time and careful observation and patience are very important. Some cells have a long crisis period, when almost all the cells appear to have died. At this stage the medium must be changed regularly, but not too frequently, with as little disturbance as possible. Eventually a few cells start to grow and form the cell line. When the cells begin to grow, caution must be taken not to split the cells too soon. Tumour cells can become overgrown by fibroblast cells which can be removed either using a rubber policeman or selective trypsinization. Serum-free medium can help to prevent fibroblast cell overgrowth.

.....

Protocol 32 Passaging cells

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.

Materials

- Versene (0.2 g l⁻¹) in isotonicly buffered saline
- trypsin/EDTA solution (0.5 g l⁻¹/0.2 g l⁻¹, Gibco in modified Puck’s solution).

Method

- 1 Pipette off growth medium, transfer to capped centrifuge tube and save, unless cell debris is present.
- 2 Wash/incubate monolayer with Versene (0.2 g l⁻¹) in isotonicly buffered saline. (For mesenchymal cells, a wash with Versene is usually sufficient; however, for epithelial cells it is advisable to incubate the cells in Versene for up to 15 min at 37 °C.)
- 3 Remove Versene and discard.
- 4 Incubate cells with trypsin/EDTA solution (0.5 g l⁻¹, 0.2 g l⁻¹ Gibco) in modified Puck’s solution at 37°C for approximately 3 min for

mesenchymal cells and up to 15 min for epithelial cells. In mixed populations of cell types, differential trypsinization can be used to enrich for one type of cell by discarding the less or more adherent cell type. The cells round up as they become detached and the trypsinization time can be monitored with an inverted microscope. Over-trypsinization can harm the cells, so incubation should be stopped when the cells are fully rounded up rather than when they become detached.

- 5 Loosen cells by tapping flask and resuspend in saved growth medium.
- 6 Return to centrifuge tube and spin at 300 g for 5 min.
- 7 Pipette off medium, loosen pellet by flicking.
- 8 Add fresh medium and split into appropriate number of new flasks. Unlike an established cell line, short-term cultures should not be split severely as the cells will settle and grow better when plated more densely.
- 9 Gas if necessary.

Protocol 33 Harvesting by removal of adherent cells

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.

Materials

- Versene (0.2 g l⁻¹) in isotonic buffered saline
- trypsin/EDTA solution (0.5 g l⁻¹, 0.2 g l⁻¹ Gibco) in modified Puck's solution
- 0.075 M KCl
- fixative (methanol/glacial acetic acid, 3 : 1)

Method

- 1 Pipette off growth medium, transfer to capped centrifuge tube and save.
- 2 Wash/incubate monolayer with Versene (0.2 g l⁻¹) in isotonic buffered saline.
- 3 For mesenchymal cells a wash with Versene is usually sufficient, but for epithelial cells it is advisable to incubate the cells in Versene for up to 15 min at 37 °C. Remove Versene and add to saved culture medium.
- 4 Incubate cells with trypsin/EDTA in modified Puck's solution (0.5 gl⁻¹

trypsin 0.2 g l⁻¹ EDTA supplied by Gibco) at 37 °C for approximately 3 min for mesenchymal cells and up to 15 min for epithelial cells. The cells round up as they become detached and the trypsinization time can be monitored by observation with an inverted microscope. Over-trypsinization can harm the cells so incubation should be stopped when the cells are fully rounded up rather than when they become detached.

- 5 Loosen cells by tapping flask and resuspend in saved growth medium and Versene.
- 6 Return to centrifuge tube and spin at 300 g for 5 min.
- 7 Pipette off medium, loosen pellet by flicking.
- 8 Resuspend pellet in 5–10 ml prewarmed 0.075 M KCl.
- 9 Incubate at 37 °C for 5–10 min.
- 10 Centrifuge for 5 min at 300 g.
- 11 Remove supernatant and resuspend pellet dropwise in precooled methanol/glacial acetic acid (3 : 1) fixative.

.....

Protocol 34 Harvesting in situ

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III. This protocol is based on that described by Mandahl [22].

Materials

- 0.3% NaCl
- fixative (methanol/acetic acid, 3 : 1)

Method

- 1 Gently remove medium from either the dish containing a coverslip or the chamber of the slide flask. Rinse briefly in 0.3% NaCl.
- 2 Rinse the coverslip or submerge the slides (in a Coplin jar) in 0.3% NaCl. Let stand for 30 min at room temperature.
- 3 Add 20% of the volume of 0.3% NaCl used in step 2 of methanol/acetic acid (3 : 1) fixative. Let stand for 5 min.
- 4 Remove 25% of the hypotonic/fixative mixture and replace with the same volume of fresh fixative. Let stand for 5 min.
- 5 Remove 50% of the mixture and replace with the same volume of fixative. Let stand for 5 min.

- 6 Remove all of the mixture and add 50 ml fixative. Let stand for 10 min.
- 7 Exchange the fixative and let stand for 30 min. Repeat once.
- 8 Withdraw the slides and let them air-dry.

.....

Troubleshooting

Unsatisfactory metaphases

Cells which are too confluent or are growing in large clumps are likely to produce metaphases that are squashed and are surrounded by cytoplasm. Metaphases at the edge of small colonies may be satisfactory. Low numbers of dividing cells can be attributed to the timing of the harvest and/or insufficient exposure to colcemid. These may be changed to optimize the number of dividing cells. It is also important to remove the medium and hypotonic/fixative carefully to avoid dislodging the dividing cells which will be adhering less strongly to the surface.

.....

Protocol 35 Labelling genomic DNA by nick translation for CGH and hybridization for CGH

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.

Materials

- dATP, dCTP, dGTP
- Tris-HCl (pH 7.8), MgCl₂, β-mercaptoethanol, BSA (nuclease free)
- fluorescently labelled dUTP nucleotides; e.g. fluorescein-12-dUTP and rhodamine-12-dUTP (FluoroGreen, FluoroRed, Amersham, UK) or fluorescein isothiocyanate (FITC)-12-dUTP and Texas red-5-dUTP (Dupont)
- DNase I (Gibco-BRL)
- DNA polymerase I (Promega)
- 0.3 mM EDTA
- Sephadex G-50
- TES buffer: 10 mM Tris-HCl, 1 mM EDTA, 0.1% SDS at pH 8.0
See Chapter 9, Protocols 46–48 for materials for hybridization.

(a) Labelling genomic DNA by nick translation

Method

- 1** In a 1.5-ml microfuge tube on ice, mix the following:
 - (a) 1 µg DNA in 39 µl distilled water;
 - (b) 5 µl of a mixture containing 0.2 mM dATP, dCTP, dGTP, in 500 mM Tris-HCl (pH 7.8), 50 mM MgCl₂, 100 mM β-mercaptoethanol, 100 µg ml⁻¹ bovine serum albumin (nuclease free);
 - (c) 1 µl (approximately 1 nmol) fluorescently labelled dUTP nucleotides (e.g. fluorescein-12-dUTP and rhodamine-12-dUTP) or fluorescein isothiocyanate (FITC)-12-dUTP and Texas red-5-dUTP);
 - (d) DNase I (approximately 200 pg to be adjusted);
 - (e) 1 µl (approximately 10 units) DNA polymerase I.
- 2** Incubate at 15 °C for 2–3 h.
- 3** Stop the reaction by the addition of 5 µl of 0.3 mM EDTA.
- 4** Purify the labelled DNA from the unincorporated nucleotide probes by passing through a Sephadex G-50 column. Sephadex G-50 swollen in TES buffer (10 mM Tris-HCl, 1 mM EDTA, 0.1% SDS at pH 8.0) is placed in a 1-ml syringe after removal of the plunger and plugging with glass wool. The column is placed in a 15-ml centrifuge tube, washed several times with the buffer by spinning at 300 g. Finally, the probe, made up to a volume of 100 µl, is placed on top of the column and spun at 300 g. The probe is collected in a microfuge tube which is placed in the bottom of the centrifuge tube before spinning.
- 5** A 10- to 20-µl sample, either before or after passing through the column, should be run on a 1% agarose gel and the double-stranded fragments sized in the range of 500–2000 bp. The amount of DNase I in the nick translation reaction, step 1d, can be altered in order to achieve this.
- 6** A sample can also be run on an agarose gel without ethidium bromide and the intensity of fluorescence visible on the transilluminator compared to a sample known to have labelled. This gives a rough indication of the amount of labelled DNA present.

.....

Troubleshooting

Standardizing conditions

Efficient labelling is essential and therefore it is advisable that each DNA sample is checked on a gel to assess the size of the fragments. Accurate measurement of the DNA concentration helps standardize the conditions and it is important that the DNA is free from impurities. If the DNA

does not produce a smear (i.e. the molecular weight remains high) or produces weak fluorescence compared to a control sample, the DNA should be repurified and/or the DNA concentration reassessed.

.....

(b) Hybridization for CGH

Method

The denaturation of the chromosomal DNA and the hybridization protocol are essentially the same as described in Chapter 9, Protocols 46–48. The amount of each DNA cohybridized is in the region of 400 ng in 10 µl of hybridization buffer and 20 µg of Cot-1 DNA is required for suppression of repetitive elements. A longer hybridization time than for single-copy probes is recommended, namely 2–3 days. If the DNA has been directly labelled, the slides can be mounted in Citifluor with 0.1 µg ml⁻¹ DAPI, immediately after washing. If biotin- and digoxigenin-labelled DNA have been used, the detection procedures detailed in Chapter 9, Protocol 49 should be followed prior to mounting.

.....

Troubleshooting

Weak or grainy signals

The appearance of signal on the chromosomes using filters specific for individual fluorochromes should be bright and primarily even along the length of the chromosome with the centromeric region free from signal. If the signal is weak or grainy, the labelling steps should be optimized with particular attention to the size of the labelled DNA (Protocol 35a). Hybridizing more DNA to the slide should also help. If the centromeric region is painted it suggests that the Cot-1 suppression may not have been effective. In addition, the actual slides and how they are made and denatured is critical. A haze specific to the region of the chromosome spreads indicates the presence of cytoplasm and problems with harvesting and the preparation of the slides. This is discussed in Chapter 9. Varying the time or temperature of chromosomal denaturation and applying the various pretreatments indicated in Chapter 9 protocols may be of benefit. However, in our experience certain batches of slides work better in CGH experiments than other batches for no obvious reason. It is therefore recommended that several batches are made and tested.

.....

Protocol 36 Tumour imprint preparation

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.

Materials

- clean slides
- 100% ethanol
- fixative (methanol/acetic acid, 3 : 1)

Method

- 1** Wash precleaned slides in 100% ethanol and wipe dry.
- 2** Place slides on a hotplate warmed to 50 °C.
- 3** Cut a small piece off frozen tumour, returning tumour to dry ice immediately.
- 4** Lightly touch the tumour onto preheated slide in areas for hybridization.
- 5** Air-dry for a few minutes. At this stage the slides generally appear to be covered in a large amount of debris; the later fixation generally cleans them up, leaving small clumps or single nuclei.
- 6** Fix in fresh methanol/glacial acetic acid (3 : 1) for 20 min.
- 7** Replace with fresh fix for a further 20 min.
- 8** Allow to air-dry.
- 9** Tumour imprint slides are stored at –20 °C, with silica gel to keep them dehydrated.

For fresh tumour samples the tumour can be lightly touched onto clean glass slides at room temperature and placed immediately in cold methanol for 20 min, being careful not to let the slide dry out. Such slides are then fixed in fresh methanol/glacial acetic acid (3:1) for 20 min. The slides are then dehydrated in 70%, 95% and 100% ethanol for 2 min each and air-dried.

.....

Protocol 37 Pretreatment of tumour imprints prior to hybridization

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.

Materials

- L15 medium
- type H collagenase blend (Sigma)
- Dulbecco's PBS with 50 mM MgCl₂
- 70%, 95%, 100% ethanol

Method

- 1 Prior to hybridization, incubate tumour imprint slides for 5 min in 50 mg of type H collagenase blend dissolved in 50 ml of L15 at 37 °C.
- 2 Rinse slides twice in Dulbecco's PBS with 50 mM MgCl₂ at room temperature for 10 min.
- 3 Dehydrate slides through 70%, 95% and 100% ethanol.

Hybridization can then be carried out using standard protocols (see Chapter 9). The temperature at which the slides are denatured is often higher (70–75 °C) and the amount of probes is often less: for example, for single-copy cosmids, 30 ng rather than the 80 ng used for hybridization to chromosomes. In order to reduce the background, higher levels of Cot-1 DNA are used, 10 µg rather than the usual 4 µg, per 10 µl hybridization mix.

Troubleshooting

Ensuring slide quality

Successful hybridization to tumour imprints depends largely on the quality of the slides. With some samples very few interphase cells attach to the slide when the tumour is touched down. Occasionally a large amount of cell debris is present on the slides which causes high background problems. Such slides can be subjected to a variety of pretreatments such as weak pepsin (100 µg ml⁻¹ dissolved in 0.01 N HCl) incubation for 20 min at 37 °C prior to collagenase digestion. The number of nuclei that should be scored depends on the type of preparations and the probes used. The possibility and the effect of normal contaminating cells should be assessed and normal controls included in experiments to indicate the false positive rate.

Acknowledgements

The authors would like to thank Professor Barry Gusterson for his helpful comments and Drs Aidan McManus and Stephen Weber-Hall for their respec-

tive contributions to the figures showing dms and the FISH analysis of the rhabdomyosarcoma case used to illustrate the approaches described. We are indebted to the Cancer Research Campaign and the Tufton Charitable Trust for their support.

References

- 1 Rabbitts, T.H. (1994) Chromosomal translocations in human cancer. *Nature* **372**, 143–149.
- 2 Mitelman, F. (1994) *Catalog of Chromosome Aberrations in Cancer* (5th edn, Wiley-Liss, New York).
- 3 du Manoir, S., Speicher, M.R., Joos, S. *et al.* (1993) Detection of complete and partial chromosome gains and losses by comparative genomic *in situ* hybridization. *Hum. Genet.* **90**, 590–610.
- 4 Kallioniemi, O.-P., Kallioniemi, A., Suudar, D. *et al.* (1993) Comparative genomic hybridization: a rapid new method for detecting and mapping DNA amplification in tumours. *Cancer Biol.* **4**, 41–46.
- 5 Gray, J.W. & Pinkel, D. (1992) Molecular cytogenetics in human cancer diagnosis. *Cancer* **69**, 1563–1542.
- 6 McManus, A.P., Gusterson, B.A., Pinkerton, C.R. & Shipley, J.M. (1995) Diagnosis of Ewing’s sarcoma and related tumours by detection of chromosome 22q12 translocations using fluorescence *in situ* hybridization on tumour touch imprints. *J. Pathol.* **176**, 137–142.
- 7 Biegel, J.A., Rorke, L.B., Parker, R.J. *et al.* (1989) Isochromosome 17q in primitive neuroectodermal tumors of the central nervous system. *Genes Chrom. Cancer* **1**, 139–147.
- 8 Pandis, N., Heim, S., Bardi, G. *et al.* (1993) Chromosome analysis of 20 breast carcinomas: Cytogenetic multiclonality and karyotypic–pathologic correlations. *Genes Chrom. Cancer* **6**, 51–57.
- 9 Birdsall, S.H., Summersgill, B.M., Egan, M. *et al.* (1995) Additional copies of 1q in sequential samples from a phyllodes tumor of breast. *Cancer Genet. Cytogenet.* **83**, 111–114.
- 10 Sambrook, J., Fritsch, E.F. & Maniatis, T. (eds) (1989) *Molecular Cloning. A Laboratory Manual*, 2nd edn. 16.30–16.55 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
- 11 Birdsall, S.H., McLennan, K.A. & Gusterson, B.A. (1992) t(6;12)(q23;q13) and t(10;16)(q22;p11) in a phyllodes tumor of breast. *Cancer Genet. Cytogenet.* **60**, 74–77.
- 12 Joos, S., Scherthan, H., Speicher, M.R. *et al.* (1993) Detection of amplified DNA sequences by reverse chromosome painting using genomic tumour DNA as probe. *Hum. Genet.* **90**, 584–589.
- 13 Telenius, H., Carter, N.P., Bebb, C.E. *et al.* (1992) Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. *Genomics* **13**, 718–725.
- 14 Shipley, J., Weber-Hall, S. & Birdsall, S. (1996) Loss of the chromosomal region 5q11–q31 in the myeloid cell line HL-60; characterization by comparative genomic hybridization and fluorescence *in situ* hybridization. *Genes Chrom. Cancer* **15**, 182–186.
- 15 Weber-Hall, S., McManus, A., Anderson, J. *et al.* (1996) Novel formation and amplification of the PAX7-FKHR fusion gene in a case of alveolar rhabdomyosarcoma. *Genes Chrom. Cancer* **17**, 7–13.
- 16 Kallioniemi, O.-P., Kallioniemi, A., Piper, J. *et al.* (1994) Optimizing comparative genomic hybridization for the analysis of DNA sequence copy number changes in solid tumours. *Genes Chrom. Cancer* **10**, 231–243.
- 17 Zitzelsberger, H., Szucs, S., Weier, H. *et al.* (1994) Numerical abnormalities of chromosome 7 in human prostate cancer detected by fluorescence *in situ* hybridization (FISH) on paraffin-embedded tissue sections with centromere-specific DNA probes. *J. Pathol.* **172**, 325–335.
- 18 Hyytinen, E., Visakorpi, T., Kallioniemi, A. *et al.* (1994) Improved technique for analysis of formalin-fixed paraffin-embedded tumors by fluorescence *in situ* hybridization. *Cytometry* **16**, 93–99.
- 19 Shipley, J., Crew, J., Birdsall, S. *et al.* (1996) Interphase fluorescence *in situ* hybridization and reverse transcription polymerase chain reaction as a diagnostic aid in synovial sarcoma. *Am. J. Pathol.* **148**, 559–567.
- 20 McManus, A.P., O’Reilly, M.J., Pritchard-Jones, K. *et al.* (1996) Interphase fluorescence *in situ* hybridization detection of t(2;13)(q35;q14) in alveolar rhabdomyosarcoma—a diagnostic tool in minimally invasive biopsies. *J. Pathol.* **178**, 410–414.
- 21 Pandis, N., Bardi, G. & Heim, S. (1994) Interrelationship between methodological choices and conceptual models in solid tumor cytogenetics. *Cancer Genet. Cytogenet.* **76**, 77–84.
- 22 Mandahl, N. (1992) Methods in solid tumour cytogenetics. In: *Human Cytogenetics, A Practical Approach*. Volume II. *Malignancy and Acquired Abnormalities*, 2nd edn (Rooney, D.E. & Czepulkowski, B.H.S., eds) 155–187 (IRL Press, Oxford).
- 23 Pandis, N., Heim, S., Bardi, G. *et al.* (1992) Improved technique for short-term culture and cytogenetic analysis of human breast cancer. *Genes Chrom. Cancer* **5**, 14–20.
- 24 Band, V. & Sager, R. (1988) Distinctive traits of normal and tumor-derived human mammary epithelial cells expressed in a medium that supports long-term growth of both cell types. *Proc. Natl Acad. Sci. USA* **86**, 1249–1253.
- 25 Lechner, J.F. & LaVeck, M.A. (1985) A serum-free method for culturing normal human bronchial epithelial cells at clonal density. *J. Tissue Cult. Methodol.* **9**, 43–48.

Fluorescence in situ hybridization

Gabriele Senger & Denise Sheer*

Institut für Humangenetik und Anthropologie, Kollegiengasse 10, D-07740 Jena, Germany

*Human Cytogenetics Laboratory, ICRF, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK

Posthybridization washes, 233

Protocol 49

Detection of hybridized probes, 234

Troubleshooting

No signal visible, 236

High background staining of chromosomes and nuclei, 236

High background staining also between chromosomes and nuclei, 237

The chromosomes are fuzzy and swollen, 237

References, 237

9.1 Introduction

Since its development over 20 years ago [1,2] *in situ* hybridization of specific nucleic acid sequences to chromosomes as a means of physically locating the positions of genes and other markers has evolved into a highly effective and rapid technique for use in gene mapping and cytogenetic analysis. At first, the nucleic acid probes were labelled with radioisotopes, and a large number of DNA probes were successfully mapped in this way. However, the use of radioactive labels limited the wide application of the technique because:

- 1 work with radioactive material is governed by strict regulations, which confined the procedure to specially authorized laboratories;
- 2 long autoradiographic exposure times are necessary for probe detection;
- 3 the poor signal-to-noise ratio meant that several metaphases had to be analysed statistically in order to define the most likely position of the DNA probe.

Several alternative methods utilizing non-radioactive labelling were described in the early 1980s [3–7], and have now almost completely replaced radioisotopic techniques. Of these, fluorescence *in situ* hybridization (FISH) is most widely used. In this technique, the probes are labelled either directly or indirectly with various fluorochrome dyes, which fluoresce in different colours when excited by UV light (see Appendix IV, Table IV.1). The location of the probe can be seen under the epifluorescence microscope, and several different probes can be visualized in a single experiment, each one being detected with a different fluorochrome [8–11]. The linear order of probes can thus be determined on metaphase chromosomes, interphase nuclei and released chromatin [12–22]. FISH also enables numerical or structural aberrations involving several different chromosomes to be analysed in a single experiment [23,24] (see Applications box 9.1).

The most commonly used fluorochromes are fluorescein isothiocyanate (FITC), which fluoresces green, Rhodamine and Texas red, which fluoresce red, with slightly different excitation and emission maxima, and 7-amino-4-methylcoumarin-3-acetic acid (AMCA), which fluoresces blue (see also Chapters 10, 11 and 13 and Appendix IV, Table IV.1). There is also a dye (Cy5) that emits in the infrared, which is not visible by eye but can be detected with a camera using an appropriate filter set [25]. The detection of individual probes with different combinations of fluorochromes further increases the number of probes that can be distinguished [10].

A gradual improvement in hybridization and detection protocols has increased the sensitivity of

the technique sufficiently to detect single-copy sequences less than 5 kb in length [26–29]. Mapping resolution has also been greatly increased by the development of techniques for releasing chromatin fibres from interphase nuclei [18–22], which provides DNA that is greatly extended compared with that in metaphase chromosomes. FISH on free chromatin fibres gives linear probe signals and so small gaps and even overlaps between two differentially labelled probes can be seen directly (Plate 3d–f).

The basic steps of a FISH protocol are:

- 1 preparation and fixation of the target (metaphase chromosomes, interphase nuclei or released chromatin) on a glass slide;
- 2 denaturation of the target DNA;
- 3 labelling of probe DNA with an appropriate reporter molecule;
- 4 incubation of the denatured probe with the target; this leads to annealing of complementary sequences;
- 5 removal of non-hybridized probe by washing and detection of the hybridized probe using avidin for biotin-labelled probes or antibodies coupled to a fluorochrome.

9.2 Slide preparation

9.2.1 Metaphase chromosomes

Routine techniques are used for the preparation of metaphase spreads as described in Protocol 13. These include hypotonic treatment of the cells and fixation in methanol/acetic acid, 3:1 (see also Protocol 39). For *in situ* hybridization it is, however, important to achieve cytoplasm-free chromosome

***In situ* hybridization is used to:**

- map probes on chromosomes [57,58] (see Plate 3a,b)
- order probes on metaphase chromosomes, interphase nuclei and released chromatin [12–22] (see Plate 3c–f)
- analyse the timing of replication of individual genes [59,60]
- investigate the three-dimensional architecture of chromatin in interphase nuclei [61]
- monitor the human chromosome content of human–rodent cell hybrid cell lines [62]
- study gene expression [63],

and in a wide range of applications to detect and analyse:

- constitutional and somatic chromosome aberrations [23,64]
- gene amplification in tumour cells [65]

Applications box 9.1

preparations. Cytoplasm surrounding the target DNA will reduce access of the labelled probe and detection reagents to the DNA and increase background staining.

Slides can be stored desiccated at -20°C for several months. DNA damage caused by acetic acid [30] is minimized by washing the slides in 70% ethanol followed by 95% and 100% ethanol before long-term storage (see also Chapter 11 on ways of avoiding DNA damage).

In order to map a probe precisely to a chromosomal region, the chromosomes must be banded before hybridization, so that chromosomes and chromosomal regions can be identified. Several banding methods can be used for FISH mapping. Protocol 38 gives a treatment for pre-banding with Wright's stain; treatment with Wright's stain usually results in a high-quality banding pattern.

Owing to high light reflection of Wright's stain or Giemsa it is not possible to visualize the banding pattern together with the fluorescence signals of a hybridized probe, and the slide needs to be destained before FISH. This rather laborious procedure of taking photographs of each metaphase twice (before and after FISH) can be avoided by using fluorescence-banding techniques that allow one to visualize the banding pattern and probe signal either simultaneously or successively by changing the fluorescence filter set. Staining of chromosomes with DAPI [31] after *in situ* hybridization and probe detection results in a Q-banding pattern (see Chapter 7) of sufficient quality for the identification of individual chromosomes. However, as small bands are not visible (Plate 3b), precise mapping of probes to subbands is not possible unless digital enhancement is used (Section 13.5).

Excellent banding can be achieved by incorporating 5-bromo-2-deoxyuridine (BrdU) into chromosomal DNA during either early or late S-phase (Plate 3a). After FISH and probe detection with a red fluorochrome, the incorporated BrdU is detected with an antibody conjugated to FITC (fluorescing green) [13,32]. As positive Giemsa (G)-bands contain late replicating DNA and negative G-bands (=positive reverse (R)-bands) early replicating DNA, this results in a G- or R-type banding pattern dependent on the timing of BrdU incorporation into the DNA during the S-phase. Protocols 39 and 40 for replication banding are routinely used in our laboratory and work well with lymphocyte cultures.

Protocols 39 and 40 in combination with replication banding using the 'fluorescence plus Giemsa' (FPG) method [33,34] result in an opposite banding pattern. Here the negatively stained bands are

the regions where thymidine has been replaced with BrdU. When fluorescein-labelled anti-BrdU antibodies (anti-BrdU-FITC) are used as the detection reagent, however, the BrdU-containing regions represent the positively stained bands. The antibody treatment for obtaining this banding pattern is described in Protocol 49b.

Instead of using fluorescein-conjugated antibodies, counterstaining of the chromosomes with propidium iodide after FISH also leads to an R-banding pattern when BrdU has been incorporated during the first half of the S-phase [35]. The quality of this banding pattern is rather variable and in our hands not as reliable as the anti-BrdU antibody technique. However, it has the advantage that the probe signals and banding pattern are visible simultaneously with a conventional fluorescence filter block for FITC (e.g. Zeiss Filter set 09).

R-banding without previous incorporation of BrdU into the chromosomal DNA can be obtained by hybridizing with labelled Alu sequences, as Alu repeats are concentrated in the negative G-bands. This method, named *in situ* hybridization banding (ISHB) [36], results in a high-quality banding pattern when Alu-PCR products generated with a single primer (No. 517: CGACCTCGAGATCT(C/T)(G/A)GCTCA CTGCAA) are used [37]. Simultaneously hybridized Alu sequences (labelled with digoxigenin) and probe DNA (labelled with biotin) can then be detected in two different colours as described in Protocol 49c.

9.2.2 Interphase nuclei

Chromatin in interphase nuclei is less condensed than in metaphase chromosomes, giving a higher mapping resolution. The signal patterns obtained on hybridization of probes to interphase nuclei differ depending on whether the target DNA has replicated. Before replication of DNA, each chromosome contains only one chromatid. Thus, hybridization signals are visible as single dots. After replication of DNA during the S-phase, the probe can hybridize to two chromatids per chromosome, visible as signal doublets. In experiments where two or three probes are hybridized simultaneously, these signal doublets can produce rather complicated signal patterns. For ordering of probes in interphase nuclei, it is therefore necessary to exclude S- and G₂-phase nuclei from the analysis. Lymphocytes isolated directly from a blood sample represent a pure population of cells arrested in the G₁ phase. They are therefore an ideal source for interphase mapping. White blood cells can be separated from erythrocytes by gravity as follows.

- 1 Attach a fresh needle (bent to give an acute angle) to the syringe containing 5–10 ml whole blood.
- 2 Incubate the syringe in an upright position (with the needle on top) for 1 h at room temperature until three zones are visible: erythrocytes, leukocytes and serum.
- 3 Remove most of the serum and collect the leukocytes, visible as a small grey zone (buffy layer) between the other two layers, in a 15-ml tube. Cells can then be collected immediately and prepared by conventional methods (see Chapter 7).

Monolayer cell cultures can be enriched for G1 interphase cells by culturing the cells at complete confluency for 3–4 days without changing the medium.

9.2.3 Released chromatin

The resolution of FISH mapping is further increased by using free chromatin fibres that are stretched and fixed on a glass slide. The hybridization signals on these preparations are visible as extended lines that can vary in length for any one probe across a slide. Several protocols for releasing chromatin from interphase nuclei have been published, mostly using fresh cells [18–20], but we have developed procedures that can be performed with routinely harvested and fixed cells [21,22] as illustrated in Fig. 9.1. These are given in Protocol 41.

9.3 Probes

A wide range of different types of probe has been employed for FISH experiments, including species-specific total genomic DNA and chromosome-specific probes for chromosome painting (both described in Chapter 10), probes that detect tandemly repeated sequences (alpha satellite, beta satellite and telomere probes), interspersed repetitive sequences (*small interspersed repetitive elements*, e.g. Alu sequences, and *large interspersed repetitive elements*, e.g. L1-elements), cosmids, phage clones, plasmids, cDNAs and RNA probes.

Alpha satellite probes are used to mark the centromeric regions of specific chromosomes—for example, to investigate aneuploidy in prenatal diagnosis or in tumours. As the number of centromeric regions can easily be counted in interphase nuclei, it is not even necessary to obtain metaphase spreads. This is especially advantageous in tumour cytogenetics, where the quality of metaphase spreads is too poor for conventional karyotype analysis or where it is difficult to obtain metaphases at all. Translocations or terminal deletions of specific

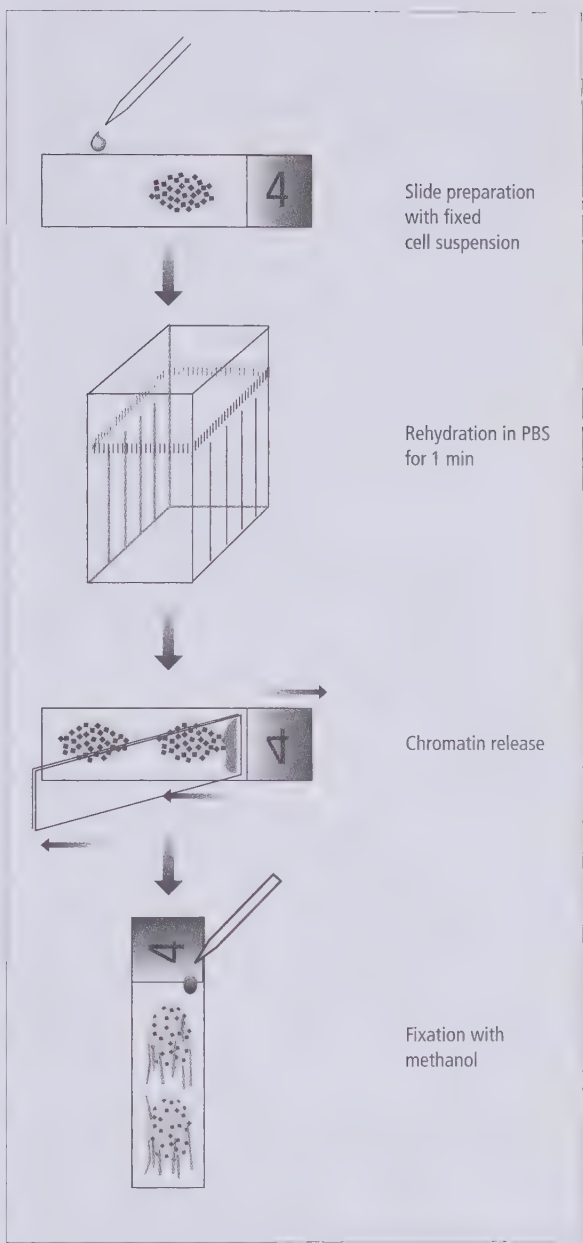


Fig. 9.1 Technique for releasing chromatin from cells fixed in methanol/acetic acid (see Section 9.2.3.).

chromosome arms can be detected using alpha satellite probes—to assist chromosome identification—together with a second single-copy probe from the same chromosome. Several such probe combinations are commercially available (Applied-Bioscience, Durham, UK) for the rapid identification of deletions characteristic of certain inherited microdeletion syndromes (e.g. cri-du-chat: del(5)(p15), Miller–Dieker: del(17)(p13.3), and Wolf–Hirschhorn: del(4)(p16.3) or tumour types (see also Chapter 10, Table 10.1).

Centromeric regions contain different monomer units, each approx. 171 bp long, which form a higher-order repeat unit. These higher-order units are tandemly repeated 100–5000 times, with a total length of up to several megabases (reviewed in ref. 38). Owing to this large target sequence, FISH signals of alpha satellite probes are very bright and are visible even without all the experimental procedures for FISH being optimized. Thus, alpha satellite probes or other repetitive probes are ideal for gaining experience with FISH in a laboratory where this technique has not been used before.

The smaller the target sequence for the hybridized probe, the more important it is to achieve the best possible conditions for performing FISH. Although probes less than 1 kb long have been detected with FISH [29], mapping of probes smaller than 2–3 kb is still rather difficult and requires a certain amount of statistical analysis. Excellent results, however, are usually obtained with lambda probes (10–20 kb), cosmids (approx. 40 kb) and yeast artificial chromosomes (YACs; 100 kb–1 Mb). In addition to unique sequences, most of these probes contain repetitive sequences, which need to be suppressed to avoid hybridization to innumerable targets throughout the genome. The suppression technique has been described as chromosomal *in situ* suppression (CISS) hybridization (also known as competitive *in situ* hybridization) [39,40] (see Chapter 10, Protocol 52). Repetitive sequences within the labelled probe can be competed either with unlabelled total human DNA or with commercially available Cot-1 DNA (BRL). We prefer Cot-1 DNA, as repetitive sequences are competed very effectively without reducing the hybridization efficiency.

The success of FISH depends to a great extent on the quality of the probe DNA. Insufficiently purified probe DNA is one of the main reasons for unsuccessful FISH results, especially when positive results with other probes exclude other factors such as inappropriate quality of one or more reagents. Minipreps of lambda DNA, plasmids and cosmids can be used after phenol/chloroform purification [41]. Purification through a column (Promega, Qiagen) or CsCl gradient is recommended for large-scale DNA preparations (see Chapter 15). In general it is not necessary to separate the insert from the vector DNA.

YACs are invaluable for cloning long fragments of genomic DNA. FISH has become an important tool for rapid and precise mapping of YACs on chromosomes and for the detection of chimaeric probes, which are present in various percentages in each YAC library. YAC DNA for FISH analysis can be obtained by several different methods. Protocol 42

for isolating the DNA from agarose plugs is currently used in our laboratory with good results in FISH experiments. The preparation of YACs in agarose blocks is described in Chapter 15 (Protocol 81).

Total yeast DNA isolated in this way can be used directly as a probe in FISH. However, compared with cosmids, a greater amount of DNA is necessary (see Protocol 47) as the proportion of probe-specific DNA in most YACs is less than 5%. Protocol 43 describes the amplification of human-specific sequences from YACs by Alu-PCR using primer AGK34 to detect Alu sequences. Other protocols have been described for amplification of human-specific sequences from YACs using various regions of the Alu consensus sequence as primer [25,42]. The fact that these methods depend on the presence and adequate spacing of Alu elements does not seem to reduce the percentage of positive FISH results considerably in comparison with unamplified total yeast DNA. In rare cases, Alu-PCR may, however, prevent detection of chimaeric YACs if one of the two co-cloned fragments is not amplified adequately.

9.4 Probe labelling

9.4.1 Choice of label

For non-radioactive *in situ* hybridization, both direct and indirect labelling methods are available.

In direct procedures the probe is labelled with a fluorochrome, which allows the visualization of the hybridized probe without any additional detection reaction. FITC-, Rhodamine- and AMCA-conjugated nucleotides are obtainable from several suppliers (e.g. Amersham, Boehringer Mannheim, DuPont, Vysis) and can be used in a nick-translation or random-primed labelling reaction [43], or during amplification of DNA by PCR [44].

In indirect methods, the probe is labelled with an invisible reporter molecule (a hapten). The most commonly used haptens are biotin and digoxigenin. These haptens are detected after *in situ* hybridization by fluorescent-labelled specific antibodies or, where biotin is the hapten, by fluorochrome-conjugated avidin. Less popular are dinitrophenol-conjugated nucleotides [40] and chemical modification of probes with sulphone groups [45], acetylaminofluorene [6,7], or mercury [46,47] to make them detectable.

Different labelling methods are especially useful for experiments in which three or more probes are hybridized simultaneously and need to be detected individually by different colours. For most applications, biotin and digoxigenin are sufficient as the

primary labels. Three probes can be distinguished using only biotin and digoxigenin, by labelling one probe with biotin, the second with digoxigenin and the third with both haptens. After probe detection (e.g. biotin with avidin conjugated to Texas red, and digoxigenin with FITC-conjugated antibodies) the hybridization signal of the doubly labelled probe is visible in both colours—red and green. When visualized through a dual band-pass filter set for simultaneous visualization of red and green fluorescence, this signal appears yellow (see Plate 3c).

Direct labelling of probes with fluorochromes has the advantage that the hybridized probe can be visualized immediately after *in situ* hybridization. Because of the absence of any intermediate detection steps, which usually increase background staining with each round of signal amplification, a low background is achieved. Probe signals are, however, weaker compared with indirect methods. Antibodies against fluorochromes are available that allow signal amplification of directly labelled probes if necessary. For centromeric probes, YACs and chromosome painting in particular (see Chapter 10) direct labelling with fluorochromes is a realistic alternative to indirect techniques, although most laboratories still prefer biotin or digoxigenin labelling.

In contrast to Southern blot hybridization, the fragment size of the labelled probe is an important factor for successful *in situ* hybridization [28,48], especially when tissue sections, intact cells or interphase nuclei are used. In these cases, only small probe fragments can penetrate through the nuclear matrix and gain access to the target DNA. For probe labelling by nick translation (Section 9.4.2, Protocol 44) the concentration of DNase I can be adjusted in order to achieve a fragment size of approx. 200–300 bp (see Chapter 10). In a random-primed labelling reaction, the fragment size is influenced by the concentration of random hexamer primers. Other labelling methods—for example, incorporation of labelled nucleotides during a PCR reaction or the chemical modification of DNA with specific haptens—require fragmentation of the probe by DNase treatment or by sonication.

Several companies provide 'kits' specifically adapted for nonisotopic labelling of DNA that include digoxigenin, biotin or fluorochrome-conjugated nucleotides. Moreover, ready-labelled probes are commercially available, including the alphoid centromere repeats, chromosome-specific paints, oncogene probes and a variety of cosmid for specific genetic loci (see Chapter 10, Table 10.1). These ready-labelled probes are especially valuable for clinical diagnostic laboratories.

9.4.2 Biotin and digoxigenin labelling

In our laboratory, probe labelling is done by nick translation using the Bionick labelling kit (Gibco-BRL) as described in Protocol 44. Since this kit includes biotinylated nucleotides, probes are usually labelled with biotin unless a different hapten is necessary for simultaneous hybridization of two or more probes. The labelling reaction is performed exactly according to the instructions of the supplier. A protocol for a less expensive nick-translation labelling procedure not using a kit is given in Chapter 10 (Protocol 50).

9.4.3 Quality control of biotin and digoxigenin labelling

It is useful to test the quality of labelling of probes during the process of establishing and optimizing FISH in a laboratory or if no signals can be obtained after FISH. This can be done by colorimetric detection of different dilutions of the labelled probe spotted on a nylon filter. All reagents necessary for the detection of biotinylated DNA (streptavidin-alkaline phosphatase, nitroblue tetrazolium and 5-bromo-chloro-3-indolylphosphate) plus DNA dilution buffer (100 $\mu\text{g ml}^{-1}$ sheared herring sperm DNA in $6\times\text{SSC}$) and biotinylated control DNA are available as a kit (BluGENE, Gibco-BRL). Similar kits can be obtained from various suppliers (e.g. Boehringer: DIG Nucleic Acid Detection kit, or Oncor: Sure Blot Blue). The procedure is given in Protocol 45.

9.5 In situ hybridization

Protocols 46 and 47 describe the denaturation of the chromosomal and probe DNA, respectively, and the hybridization of the probe to the chromosomal DNA. After FISH, the morphology of chromosomes is better preserved in older slides than in fresh ones. This is especially evident in banded chromosomes. For slides that need to be used less than one week after preparation, ageing can be speeded up by heat treatment before denaturation. Pretreatment of slides with RNase is usually not necessary for FISH. With some probes that produce high background staining as a result of hybridization with RNA, pretreatments described in Protocol 46, steps 1–3 might be useful. In most cases, the first three steps can be skipped. Several FISH protocols include the pretreatment with proteinase K or pepsin in order to remove residual proteins. Here, however, the optimal enzyme concentration is different for each cell type and very critical for retaining the

chromosome morphology. Therefore we try to avoid cytoplasm during the cell-harvesting procedure and omit treatment with proteinase K or pepsin.

9.6 Probe detection

Non-hybridized probe and nonspecific annealed probe DNA need to be removed before probe detection. This is described in Protocol 48. Detection steps are described in Protocol 49.

Protocol 49c is also suitable for three-colour FISH using three probes: one labelled with biotin, the second with digoxigenin and the third with biotin and digoxigenin. Another approach for obtaining three colours with this protocol is to label one probe with a Rhodamine-conjugated nucleotide (which is visible without additional detection reagents) and the other two probes with biotin and digoxigenin, respectively. The biotinylated probe is detected in blue with avidin-AMCA (diluted 1:50) instead of avidin-Texas red. The brightness and stability of the three fluorochromes decreases in the order FITC > Rhodamine/Texas red > AMCA, which might direct the choice of labelling systems for each of the three probes. The probe that gives the weakest signal should be labelled with digoxigenin and detected with FITC. As the detection of a biotinylated probe with AMCA following Protocol 49c includes one round of signal amplification, there is no obvious difference between the signal intensities of the probe directly labelled with Rhodamine and the biotinylated probe detected with AMCA. However, if one probe produces notably higher background, it is advantageous to use direct Rhodamine labelling for this probe.

In all multicolour FISH experiments, it is important to check for and exclude cross-reactivity between the antibodies used for detecting differentially labelled probes. If a highly sensitive CCD camera is available for capturing microscopic images, signal amplification should be done by image processing rather than by detecting the probe with several layers of fluorochrome-conjugated antibodies. Chapter 13 provides a review of equipment for digital fluorescence microscopy.

9.7 Analysis of probe signals

9.7.1 Microscopy equipment

Viewing of probe signals obtained with FISH requires a high-quality epifluorescence microscope (e.g. Axioskop, Zeiss) equipped with at least three different filter sets specific for FITC, Texas red/Rhodamine and DAPI/AMCA. Chapter 13 provides a review of equipment for digital fluorescence microscopy. The characteristics of filter sets (Zeiss) used in our laboratory are given in Table 9.1 (filter sets for the Nikon Optiphot microscope are given in Appendix IV, Table IV.2).

For precise mapping of probes (detected in red) on an FITC-stained chromosome banding pattern a dual band-pass filter set (Omega, Zeiss, Chroma Technology) is ideal as it enables simultaneous visualization of both fluorochromes. Red, green and blue fluorescence can be seen simultaneously with a triple band-pass filter set, which is especially useful for ordering three probes, detected with FITC, Texas red and AMCA. Alternatively, separate pictures obtained with different filter sets – each specific for only one fluorochrome – can be merged either by double and triple exposure of the same photograph or by processing digitized images with specific computer software (see Chapter 13). Both methods require precise alignment of the filter sets in order to avoid incorrect mapping due to optical shift. As a control experiment one probe detected simultaneously in two or three colours should result in entirely overlapping probe signals.

9.7.2 Mapping strategies

9.7.2.1 Localization on metaphase chromosomes

The first step in mapping new probes is their localization on banded metaphase (preferably prometaphase) chromosomes (see Case Study 9.1). If hybridization signals are consistently found on the same chromosome band and on both sister chromatids, it is usually sufficient to screen less than five metaphases in order to determine the hybridization site. Multiple hybridization sites for a probe can

Table 9.1 Filter sets used for fluorescence detection and analysis (see also Appendix IV, Tables IV.2 and IV.4).

Fluorochrome	Exciter filter	Dichroic reflector	Barrier filter	Filter set (Zeiss)
FITC + propidium iodide	BP 450–490	510	LP 515	09
Texas red/Rhodamine	BP 546	580	LP 590	15
AMCA/DAPI	G 365	395	LP 420	0

G, solid glass filter; BP, bandpass filter; LP, long-wave bandpass filter.

occur as a result of coligated DNA fragments or sequence homology between two or more genomic regions. If it is difficult to identify true probe signals because of high background and/or weak probe signals, a greater number of metaphases need to be analysed.

9.7.2.2 Ordering on metaphase chromosomes

Probes can be ordered on metaphase chromosomes by cohybridizing them in differentially labelled pairs. There must be a minimum distance of roughly 1–3 Mb between the probes in order to see the signals separated well enough for ordering [15,16]. Chromatin folding, however, may lead to an incorrect probe order and needs to be excluded by analysing several metaphases if the two signals are relatively close. At the telomeric end of a chromosome, ordering using FISH can be difficult, as the signals of telomeric probes can appear to be proximal to the visible chromosome end [15,49].

9.7.2.3 Ordering in interphase nuclei

If two probes are too close for their signals to be resolved even in prometaphase chromosomes, ordering can be performed in interphase nuclei. As ordering in relation to the telomere or centromere is not possible, a third reference probe is necessary. All three probes can be ordered in a single experiment by hybridizing them simultaneously and by detecting each probe in a different colour. This can be achieved with just two different haptens (biotin and digoxigenin) by labelling two probes with one of each hapten and the third probe with both. Thus any optical shift is instantly obvious and calculable even if no dual band-pass filter set is available for simultaneous visualization of Texas red and FITC. On the other hand, the mixed colour yellow emerges not only from a doubly labelled probe but also from partially overlapping red and green probe signals or from green signals on red background (or vice versa). Therefore, ordering of three probes, using double labelling for obtaining a third colour, is only possible with non-overlapping probe signals on low background. Another option is the detection of two probes in the same colour and the third probe in a second colour. In this system, however, up to three experiments are necessary as the probe order is only evident if the probe in the middle is labelled differently to the two flanking probes (red–green–red or green–red–green).

Chromatin folding and the analysis of a three-dimensional object on a two-dimensional picture will lead to an incorrect signal order in a proportion of nuclei. This becomes increasingly problematic with probes that are more than 1–2 Mb apart.

Mapping the TRAP gene, the human homologue of the murine CD40L gene

The chromosomal localization of a gene that is responsible for a well-characterized inherited disease can often be identified by specific cytogenetic abnormalities or by multipoint linkage analysis even if the gene itself is not cloned. Precise mapping by fluorescence *in situ* hybridization can sometimes reveal the causal relationship between a gene whose function is unknown and an inherited disorder that maps to the same locus. One example is the gene coding for a tumour necrosis factor-related activation protein (TRAP) [66–68]. The TRAP cDNA was isolated from a gt10 cDNA library generated from T-cells stimulated with the mitogen phorbol myristate acetate (PMA) and Ca^{2+} ionophore A23187. Sequence analysis revealed similarity to tumour necrosis factor- α (TNF- α) and lymphotoxin (TNF- β). Because of its close homology to a ligand for the murine CD40 molecules (CD40L), TRAP was identified as the human homologue of the murine CD40L, which is expressed on the surface of activated T cells.

The TRAP/CD40L gene was mapped to the long arm of chromosome X in the region Xq26.3–q27 using fluorescence *in situ* hybridization with a 15-kb genomic TRAP gene probe. Simultaneous visualization of the chromosome banding pattern was obtained with the BrdU antibody technique. The hyper-IgM immunodeficiency syndrome (HIGM1) was mapped previously close to the hypoxanthine phosphoribosyl transferase (HPRT) gene in the same region (Xq26) by multipoint linkage analysis [69]. HIGM1 is a rare disorder characterized by the lack of IgG and IgA production together with a normal or increased IgM level. Taken together with the fact that the B-cell proliferation and immunoglobulin isotype switch is stimulated via CD40 in the presence of activated T-cells, these findings suggested a causal relationship between TRAP/CD40L and HIGM1. This hypothesis was finally confirmed by the demonstration that different point mutations in the TRAP/CD40L gene from several HIGM1 patients resulted in functionally defective CD40L molecules (reviewed in ref. 68).

Case Study 9.1

Ordering in interphase nuclei is also difficult when the distances between the probe in the middle and the two flanking probes are very different (e.g. < 300 kb and > 1 Mb) [16]. Here the orientation of the two closely spaced probes will be random relative to the third probe. In general, for ordering in interphase nuclei a statistical analysis is necessary in order to exclude incorrect probe ordering.

Probes can be ordered in interphase nuclei by measuring and comparing interphase distances between the signals of pairwise cohybridized probes. The average distances measured between each probe pair have been found to be strongly correlated with distances in kilobases [12–16]. By comparing the calibration curves obtained in

different studies [12–16], it is, however, evident that the relationship between interphase and kilobase distances is not consistent for all genomic regions.

9.7.2.4 Mapping on released chromatin

FISH on free chromatin fibres fixed on a glass slide enables one to map and order directly adjacent or overlapping probes. The decondensation of chromatin is variable in different areas of one slide and the length of probe signals can reach up to twice the theoretical length of the DNA double helix (3.4 Å per bp) [20–22]. In contrast to interphase nuclei, there is no relationship between measured distances and kilobase distances. But if the kilobase length of the probe is known, the measured length of a probe signal can be used as an internal ruler (individually for each signal pair) by calculating the kb per µm of probe signal. Thus, the DNA length of overlaps and small gaps between two probes can be determined [22]. As in interphase measurements, a statistical analysis is required to eliminate errors due to in-

consistent chromatin decondensation within probe signals or to shorter signals caused by DNA breaks.

Gaps between two probes are more difficult to analyse than overlaps. In the region of overlap the two differentially labelled probe signals follow the same line and it is obvious which signals are hybridized to the same DNA fibre. In the analysis of nonoverlapping probes, the selection of probe signals that are accepted as signal pairs is somewhat subjective. Therefore it is always necessary to confirm the close proximity of probes in interphase nuclei before using chromatin release techniques.

When using FISH techniques for mapping and ordering probes or for clinical diagnosis one has to be sure that the results obtained are not a product of selection bias. Control experiments with a known result (but not known to the person performing the experiment) are therefore very important for monitoring the reliability of the results obtained with FISH for each particular application.

Protocol 38 Pre-banding using Wright’s stain

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- Wright’s stain (Gurr)
- Sørensen buffer
- SSC buffer

Method

- 1 Prepare the staining solution from 1 part Wright’s stock solution and 3 parts 50% Sørensen buffer, and pass through a Whatman filter.
- 2 Incubate the slide in 2×SSC at 60 °C for 5–10 min. The optimum time varies with the age of a slide and needs to be increased for older slides.
- 3 Stain the slide for 1–2 min, rinse with water and air-dry.
- 4 Mount the slide in one drop of water and take photographs from a selection of well-spread and banded metaphases. As these metaphases have to be relocated after *in situ* hybridization, it is important to make a note of the exact position of each metaphase.
- 5 Destain the slide with methanol and air-dry.

Protocol 39 BrdU-incorporation during late S-phase for replication G-banding

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- RPMI 1640 medium (Seromed)
- FCS (Seromed)
- L-glutamine (Seromed)
- penicillin/streptomycin (Seromed)
- phytohaemagglutinin (Wellcome Diagnostics)
- 5-bromo-2-deoxyuridine (BrdU) (Sigma)
- 5-fluorodeoxyuridine (FdU) (Sigma)
- thymidine (Sigma)
- uridine (Sigma)

Method

- 1 Prepare a 10-ml lymphocyte culture with:
 - 9.5 ml RPMI medium containing 10% FCS, 2.4 mM L-glutamine, and 100 U ml⁻¹ penicillin, 100 µg ml⁻¹ streptomycin;
 - 0.1 ml phytohaemagglutinin (9 mg ml⁻¹);
 - 0.5 ml whole blood (heparinized with 50 U ml⁻¹ sodium heparin); and incubate for 72 h at 37 °C.
- 2 For synchronization of the cells within the S-phase add 100 µl 5-fluorodeoxyuridine (0.05 mM) (FdU; end conc., 5×10^{-7} M) and 100 µl uridine (0.12 mg ml⁻¹) (end conc., 1.2 µg ml⁻¹) and incubate overnight (16–20 h).
- 3 Add 15 µl BrdU (20 mg ml⁻¹) (end conc., 30 µg ml⁻¹) and incubate for a further 7 h before harvesting.
- 4 Centrifuge at 1500 r.p.m. for 5 min and remove the supernatant with a pipette, leaving about 0.5 ml above the pellet.
- 5 Resuspend the cells in 10 ml hypotonic solution (0.075 M KCl), prewarmed to 37 °C and incubate for 10 min at room temperature.
- 6 Centrifuge at 1500 r.p.m. for 5 min.
- 7 Remove the supernatant with a pipette, resuspend the cells in the remaining drop of hypotonic solution and take the cell suspension into the pipette.
- 8 Fill the tube with 10 ml ice-cold fixative (methanol/glacial acetic acid, 3:1) and quickly immerse the cells into the fixative. Mix by inverting the tube.
- 9 Incubate for 20–30 min on ice.

- 10 Wash the cells 2–3 times with fixative.
- 11 Drop the cells onto clean wet slides, air-dry and remove traces of fixative by washing the slides in 70%, 95% and 100% ethanol for approximately 3 min each.
- 12 Air-dry and store the slides desiccated at –20 °C.

FdU is converted into its phosphorylated nucleotide, FdUMP, which inhibits thymidylate synthetase, the enzyme necessary for thymidine synthesis. Uridine is added to avoid incorporation of fluorouridylate, which can be formed from FdU, into RNA [50]. Despite the FdU block, the cells are able to synthesize sufficient amounts of thymidine to progress through the early S-phase but most cells do not go beyond this stage [51]. Addition of BrdU then leads to release of the block on DNA synthesis without the need to wash the cells.

.....

Protocol 40

BrdU incorporation during early S-phase for replication R-banding

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- RPMI 1640 medium (Seromed)
- FCS (Seromed)
- penicillin/streptomycin (Seromed)
- phytohaemagglutinin (Wellcome Diagnostics)
- 5-bromo-2-deoxyuridine (BrdU) (Sigma)
- thymidine (Sigma)

Method

- 1 Prepare a 10-ml lymphocyte cell culture (see Protocol 39) and incubate for 72 h.
- 2 Synchronize the cells within the S-phase by adding 100 µl BrdU (20 mg ml⁻¹) (end conc., 200 µg ml⁻¹) and incubate overnight (16–20 h).
- 3 Centrifuge for 5 min at 1500 r.p.m. and wash the cells once in medium without serum (prewarmed to 37 °C).
- 4 Pellet the cells again and resuspend them in 10 ml medium containing 10% serum and 2.5 µg ml⁻¹ deoxythymidine.
- 5 Incubate for a further 6½ h before harvesting.
- 6 Continue as described above in the protocol for replication G-banding.

During and after treatment with BrdU, exposure of cells and meta-phase spreads to bright light should be avoided whenever possible. Each slide can be briefly inspected using phase-contrast microscopy in order to locate the best areas for FISH.

These protocols are also suitable for cells other than lymphocytes, but it may be necessary to determine the optimal time between release of the block and harvesting. Treatment with colcemid is usually not necessary, as colcemid leads to accumulation of relatively short chromosomes but does not increase the yield of prometaphase chromosomes, and has no effect on the quality of metaphase spreads [52].

Protocol 41 The release of chromatin from interphase nuclei

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- formamide (FSA Laboratory Supplies)
- PBS
- SSC buffer

Method

- 1 Place one drop of fixed cell suspension on each half of a clean wet glass slide.
- 2 Before the fixative starts to evaporate, transfer the slide into a Coplin jar filled with PBS in order to rehydrate the cells. As the cells are only loosely attached to the slide, avoid any agitation of the slide during this incubation step.
- 3 After 1 min remove the slide and drain it on a paper towel, without allowing the slide to dry.
- 4 Place 100 µl NaOH (0.07 M) mixed with ethanol^a in the ratio 5 : 2 on one end of a long coverslip (24 × 60 mm) and move this solution carefully over the cells. The best way is to hold the slide with the cells upside down at a small angle to the coverslip as shown in Fig. 9.1.
- 5 The released chromatin fibres are then fixed by rinsing the slide with methanol. Loss of chromatin during this step can be avoided by starting with a minimal amount of methanol added on one end of the slide, which is held horizontally. Viscous fluid dropping off the slide indicates loss of chromatin. Try to keep the methanol on the slide for as long as possible, until the chromatin, visible as a gelatinous bulk, is attached to the slide.

^aSeventy per cent formamide in 2 × SSC (pH 7.0), can be used instead of NaOH/ethanol for releasing chromatin. Using this procedure, the borders of most nuclei are still visible, which is especially useful when it is necessary to identify hybridization signals derived from one nucleus. Chromatin released with formamide is generally less decondensed than with NaOH, as can be seen from comparisons of signal lengths. In contrast to the NaOH method, where cells that have been stored in fixative for several months can be used, formamide release requires that cells are stored in fixative for no longer than a few days.

- 6 Air-dry the slide and check its quality using phase contrast microscopy. Released chromatin is visible as a network of fibres across the slide. These slides can be used the same day for FISH or stored at -20°C .

.....

Protocol 42 DNA isolation from agarose plugs

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- agarose plugs containing YAC DNA
- agarase (New England Biolabs)
- *Hind*III-digested lambda DNA (Gibco BRL)
- TE buffer, TAE buffer
- buffer-saturated phenol/chloroform/iso-amylalcohol (24 : 23 : 1)
- 3 M sodium acetate, pH 5.6
- ethanol
- Eppendorf tubes

Method

- 1 Transfer two or three plugs into an 1.5-ml Eppendorf tube.
- 2 Add 200 μl TE and melt the agarose in a 68°C water bath for 10 min.
- 3 Cool to 37°C for 5 min, add 2–3 units agarase and incubate for 1 h at 37°C .
- 4 Place tube on ice for 5 min.
- 5 Purify the DNA twice with an equal volume of buffer-saturated phenol/chloroform/iso-amylalcohol (24 : 23 : 1) (mix by vortexing) and once with chloroform.
- 6 Add $\frac{1}{10}$ vol. 3 M sodium acetate (pH 5.6), and 2 vols ethanol and freeze on dry ice for 10 min.
- 7 Pellet the precipitated DNA in an Eppendorf centrifuge at maximum speed for 15 min at 4°C .
- 8 Wash the pellet once in 70% ethanol, dry and resuspend in 20 μl double-distilled water.
- 9 Check the DNA concentration on a 1% agarose gel in $1\times\text{TAE}$ by using 2 μl of the YAC DNA run against 500 ng of lambda DNA digested with *Hind*III. The quantity of DNA can be estimated by comparing the intensity of the band from the YAC DNA with the lambda standard. Table 9.2 shows the amount of DNA in each lambda fragment.

Table 9.2 DNA content in the restriction fragments of lambda DNA digested with *Hind*III.

Fragment size (kb)	% of total DNA	ng (500 ng total)
23.13	47.7	238.5
9.42	19.4	97.0
6.56	13.5	67.5
4.36	9.0	45.0
2.32	4.8	24.0
2.02	4.2	21
0.56	1.2	6
0.12	0.2	1

.....

Protocol 43 Alu-PCR with primer AGK34 [25]

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Best results are obtained with purified YAC DNA as template. Alternatively, the use of molten agar blocks or even yeast cells without further DNA purification has been described [25,42] (see Chapter 15), but the yield of amplified DNA is less reproducible.

A protocol for Alu-PCR on flow-sorted human chromosomes using primer BK-33, to produce chromosome paints, is given in Chapter 10, Protocol 54.

Materials

- YAC DNA
- Taq polymerase (Perkin Elmer/Cetus)
- 10×PCR buffer (Perkin Elmer/Cetus): 15 mm MgCl₂, 100 mm Tris-HCl, 500 mm KCl, 0.01% gelatine
- dATP, dCTP, dGTP, dTTP (Boehringer)
- light mineral oil (Sigma)
- 3 M sodium acetate, pH 5.6
- ethanol
- Eppendorf tubes

Method

- 1 For a 50-µl PCR reaction mix:
 - 100 ng YAC DNA;
 - 0.5 µM primer AGK34;
[5'GAGCCGAGATCG(C/T)GCCACTGCACTCCAGCCTGGG3'];
 - 5 µl 10×PCR buffer;
 - 200 µM of each of the four dNTPs;
 - 2 units of Taq polymerase (Perkin-Elmer/Cetus);and overlay with 50 µl light mineral oil.
- 2 After 3 min denaturation at 94 °C amplify the DNA in 30 cycles with:
1 min at 94 °C;

45 s at 55 °C;
5 min at 68 °C.

- 3 Transfer the PCR-reaction mix into a fresh Eppendorf tube.
- 4 Precipitate the DNA with $\frac{1}{10}$ vol. 3 M sodium acetate (pH 5.6), and 2 vols ethanol, and resuspend in 20 µl double-distilled water.
- 5 Check the DNA concentration on a 1.2% agarose gel and use 1 µg for labelling by nick translation.

.....

Protocol 44

DNA labelling with biotin and digoxigenin by nick translation using a commercial kit

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- Bionick kit (Gibco-BRL)
- digoxigenin-11-dUTP (Boehringer)
- biotin-14-dATP (Gibco-BRL)
- BSA (nuclease-free) (Boehringer)
- β-mercaptoethanol (Sigma)
- dATP, dCTP, dGTP, dTTP (Boehringer)
- Sephadex G-50 (Pharmacia)
- *Escherichia coli* tRNA (Boehringer)
- salmon testes DNA type III (Sigma)
- TE buffer
- microcentrifuge tubes
- Pasteur pipette

Method

- 1 Mix in a 1.5-ml microcentrifuge tube placed on ice:
 - 1 µg probe DNA (cosmid, phage or YAC) with:
 - sterile, distilled H₂O to 40 µl;
 - 5 µl 10×dNTP/buffer mix;
 - 5 µl enzyme mix.

For labelling with digoxigenin or for double labelling with biotin and digoxigenin, exchange the 10×dNTP-mix for one of the two mixtures described in Tables 9.3 and 9.4.

- 2 Incubate at 16 °C for 60 min.
- 3 Add 5 µl stop buffer.

Table 9.3 Preparation of 10× dNTP containing digoxigenin-dUTP.

Component	Volume for 100 µl	Concentration in 10×
10 mM dCTP	2 µl	0.2 mM
10 mM dGTP	2 µl	0.2 mM
10 mM dATP	2 µl	0.2 mM
10 mM dTTP	1 µl	0.1 mM
1 mM digoxigenin-dUTP	10 µl	0.1 mM
1 M Tris-HCl pH 7.8	50 µl	500 mM
1 M MgCl ₂	5 µl	50 mM
10 M β-mercaptoethanol ^a	1 µl	100 mM
20 mg ml ⁻¹ nuclease-free BSA	0.5 µl	100 µg ml ⁻¹
H ₂ O	26.5 µl	

^a100% β-mercaptoethanol = 14.33 M.

Table 9.4 Preparation of 10× dNTP with digoxigenin-dUTP and biotin-14-dATP.

Component	Volume for 100 µl	Concentration in 10×
10 mM dCTP	2 µl	0.2 mM
10 mM dGTP	2 µl	0.2 mM
10 mM dATP	1 µl	0.1 mM
10 mM dTTP	1 µl	0.1 mM
1 mM digoxigenin-dUTP	10 µl	0.1 mM
0.4 mM biotin-14-dATP	25 µl	0.1 mM
1 M Tris-HCl pH 7.8	50 µl	500 mM
1 M MgCl ₂	5 µl	50 mM
10 M β-mercaptoethanol ^a	1 µl	100 mM
20 mg ml ⁻¹ nuclease-free BSA	0.5 µl	100 µg ml ⁻¹
H ₂ O	2.5 µl	

^a100% β-mercaptoethanol = 14.33 M.

4 Unincorporated nucleotides are removed by passing the reaction through a Sephadex G-50 column. This column is prepared in a 145-mm Pasteur pipette that is plugged with sterile glass wool and placed in a 1.5-ml microcentrifuge tube. Sephadex G-50, swollen in TE buffer (10 mM Tris-HCl, 1 mM Na₂-EDTA, pH 8), is filled into the pipette (avoid air bubbles) up to approx. 10 mm below the top. Wash the column once with 500 µl TE buffer. Add the probe mix (= 55 µl) and allow to enter into the gel. Add 545 µl TE buffer and discard the eluate. Place the column in a fresh microcentrifuge tube, add another 600 µl TE and collect the eluate, which contains the labelled probe.

The exact volumes needed for collecting the right fraction that contains the labelled probe DNA may be monitored by adding 0.5% blue dextran to the labelling mixture before loading on the Sephadex column. Both colour and probe DNA are eluted in the same fraction.

- 5** To the fraction with labelled DNA add:
- 0.1 vol. 3 M sodium acetate, pH 5.6;
 - 5 µl *E. coli* tRNA (10 mg ml⁻¹); and
 - 5 µl salmon sperm DNA (10 mg ml⁻¹ sonicated to approx. 500 bp long).
- 6** Mix and transfer half of the volume into a fresh Eppendorf tube.

- 7 Precipitate the DNA in each tube with 2 vols ethanol.
- 8 Mix and freeze in dry ice for 10 min or at -20°C overnight.
- 9 Pellet the precipitated DNA in an Eppendorf centrifuge at maximum speed 15 min at 4°C .
- 10 Discard the supernatant and dry the pellet in a spin vac.
- 11 Resuspend the DNA of both tubes in a total volume of $50\text{ }\mu\text{l}$ H_2O and combine the two tubes. In case of YAC DNA that has been labelled together with the genomic yeast DNA, resuspend the pellet of each tube in $10\text{ }\mu\text{l}$ H_2O and use one tube (= 500 ng DNA) for one experiment.
- 12 Stored frozen, the labelled DNA is stable over months. Under sterile conditions it can be thawed and refrozen several times without loss of quality.

.....

Protocol 45 Quality control of biotin and digoxigenin labelling

Materials

- BluGENE kit (Gibco-BRL)
- alkaline phosphatase-linked anti-digoxigenin antibodies (Boehringer)
- buffer 1: 100 mM Tris-HCl, 150 mM NaCl, pH 7.5
- buffer 2: 100 mM Tris-HCl, 100 mM NaCl, 50 mM MgCl_2 , pH 9.5
- blocking buffer (5% fat-free dried milk in buffer 1)
- streptavidin-alkaline phosphatase
- nitroblue tetrazolium
- 5-bromo-4-chloro-3-indolylphosphate (50 mg ml^{-1} in dimethylformamide)
- TE

Method

- 1 Dilute the labelled DNA ($1\text{ }\mu\text{l} = 20\text{ ng}$) with DNA dilution buffer to 100, 10, 5, 2, 1 and $0\text{ }\mu\text{g }\mu\text{l}^{-1}$. Prepare the same concentrations from labelled control DNA provided with the kit or from an own-labelled probe successfully used in previous experiments.
- 2 Spot $1\text{ }\mu\text{l}$ of each dilution step (own labelled and control DNA) on a small piece of nylon filter.
- 3 Bake the filter for 30 min at 80°C .
- 4 Place the filter in a 50-ml polypropylene tube and wash briefly in buffer 1 at room temperature.
- 5 Decant buffer 1 and add 5 ml of blocking buffer (5% fat-free dried milk (Marvel) in buffer 1). Incubate for 30 min at 37°C .

- 6 Decant the blocking buffer and add 2–3 ml of a freshly prepared solution of streptavidin–alkaline phosphatase (SA–AP) diluted to 1.0 µg ml⁻¹ in buffer 1. For detection of digoxigenin-labelled probes use sheep anti-digoxigenin antibodies conjugated to alkaline phosphatase diluted to 150 mU ml⁻¹ in buffer 1. Incubate for 30 min at 37 °C (clamped on a rotating wheel).
- 7 Wash 3×5 min in 30 ml buffer 1 at room temperature.
- 8 Equilibrate the filter for 10 min in buffer 2.
- 9 Just before use prepare the dye solution in a 15-ml polypropylene tube by gently mixing 33 µl nitroblue tetrazolium (NBT, 75 mg ml⁻¹ in 70% dimethylformamide) with 7.5 ml buffer 2 and by adding 25 µl 5-bromo-4-chloro-3-indolylphosphate (BCIP, 50 mg ml⁻¹ in dimethylformamide). Mix by inverting the tube.
- 10 Incubate the filter in the NBT/BCIP solution at room temperature for 30 min to 3 h in the dark. Do not agitate during this colour reaction.
- 11 Stop the reaction by washing the filter in TE.

The probe is labelled well if 1–10 pg of probe DNA is visible.

.....

Protocol 46 Denaturation of chromosomal DNA

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- RNase (Sigma)
- formamide (FSA Laboratory Supplies)
- SSC buffer
- ethanol

Optional pretreatment

- 1 Place 100 µl RNase (100 µg ml⁻¹ in 2×SSC) on the slide, cover it with a 24×50 mm coverslip and incubate the slide for 1 h at 37 °C in a moist chamber.
- 2 Wash the slide 3×3 min in 2×SSC at room temperature.
- 3 Pass the slide through an ethanol series of 70%, 95% and absolute ethanol for 3 min each and air dry.

Routine protocol

- 4 Bake slides at 60–65 °C for 2–3 h.
- 5 Place the slide in a Coplin jar with 70% formamide, 2×SSC (pH 7.0), prewarmed to 75 °C for 3 min.
Note: This step is critical: on the one hand, sufficient denaturing of chromosomal DNA is necessary for successful *in situ* hybridization, but on the other hand, high temperatures can lead to a fuzzy chromosome structure, especially with freshly prepared slides. If more than one slide is being denatured, the temperature must not drop below 70 °C. Usually temperatures between 70 °C and 75 °C give good results.
- 6 Dehydrate the slide through an ethanol series of 70% (ice-cold), 95% and 100% ethanol for 3 min each.
- 7 Air-dry the slide.

Although denaturation in 70% formamide/2× SSC is preferred by most laboratories alternative methods have been described [53]. Equally good results can be achieved with denaturation in 0.15 M NaOH, 70% ethanol for 4 min at room temperature followed by dehydration in 70%, 95% and 100% ethanol.

Protocol 47 Denaturation and prehybridization of probe DNA and hybridization to chromosomal DNA

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Each hybridization experiment normally occupies half a slide. The concentration of labelled DNA is calculated without considering the loss of DNA during purification as being 1 µg resuspended in 50 µl TE. For hybridization of cosmid probes we use 80 ng labelled DNA (=4 µl). Slightly more DNA (100–200 ng) is used for smaller probes such as lambda clones, plasmids, or cDNAs. For probes with a highly repetitive target sequence (e.g. alphoid centromere repeats or Alu-probes) 20 ng (=1 µl) is sufficient. From YAC probes that contain the entire yeast genome in addition to the actual probe DNA, we use half of the labelled DNA (=500 ng in 10 µl), but 100 ng if the probe has been generated by Alu-PCR.

Materials

- human Cot-1 DNA (Gibco-BRL)
- hybridization mix: 2×SSC, 50% formamide, 10% dextran sulphate, 1% Tween 20
- ethanol

Method

- 1 Mix the amount of DNA required for one experiment with
 - 4 µg Cot-1 DNA (=4 µl)
 - 2 vol. 100% ethanol.
- 2 Dry the DNA in a spin vac and resuspend the pellet in 12 µl hybridization mix.
- 3 Denature the probe DNA at 75–80 °C for 3 min.
- 4 Chill the DNA on ice and spin quickly to get all the liquid down to the bottom of the tube.
- 5 Preanneal repetitive sequences by incubation at 37 °C for 30 min.
Probes that are known not to contain any repetitive sequences do not need to be prehybridized with competitor DNA and may be applied to the slide immediately after denaturation.
- 6 *Hybridization* The preannealed probe is placed on one half of the previously denatured slide and covered with a 22×22 coverslip. This is sealed with rubber cement and the slides are placed in a moist chamber and incubated at 37 °C for 24–72 h. A plastic sandwich box can be used with a sheet of paper towel moistened with water for a moist chamber.

Although it has been reported that renaturation of chromosomal DNA occurs rapidly and hybridization is essentially complete after 4 h [28], in our experience prolonged hybridization times (up to 3 days) do lead to stronger signals, which is especially advantageous for small single-copy probes (cDNA, plasmid DNA). On the other hand background can increase as well.

.....

Protocol 48 Posthybridization washes

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- formamide (FSA Laboratory Supplies)
- SSC buffer
- SSCT

Method

- 1 Prewarm 300 ml 50% formamide/2×SSC (pH 7.0), and 300 ml 2×SSC at 42 °C.

- 2 Peel off the rubber solution and carefully remove the coverslip. This can be eased by soaking the slide in the first washing solution.
- 3 Wash the slides:
 - 3×5 min in 50% formamide, 2×SSC (pH 7.0) at 42 °C;
 - 3×5 min in 2×SSC (pH 7.0) at 42 °C; and
 - 1×3 min in SSCT at room temperature.

Protocol 49 Detection of hybridized probes

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Overview

- (a) Detection of biotin with FITC
- (b) Detection of biotin with Texas red on chromosomes banded with anti-BrdU–FITC
- (c) Detection of biotin with Texas red and digoxigenin with FITC

Materials

- SSCTM
- SSCT
- labelled detection reagents as appropriate: avidin–FITC (Vector Labs), biotinylated anti-avidin (goat) (Vector Labs), avidin–Texas red (Vector Labs), anti-BrdU–FITC (Boehringer), mouse anti-digoxigenin (monoclonal) (Boehringer), FITC-labelled sheep anti-mouse Ig (Boehringer)
- Citifluor AF1 (Citifluor Ltd)
- 4,6-diamidino-2-phenylindol-2HCl (DAPI) (Sigma), or
- propidium iodide

All antibodies are diluted in SSCTM in the ratios listed below in each protocol. For every detection step, calculate 100 µl antibody solution per slide. The first ‘blocking’ step is identical for all detection protocols.

- 1 Apply 100 µl of SSCTM on each slide and cover with a 22×50 mm coverslip (avoid air bubbles). Incubate for 15 min at 37 °C in a moist chamber followed by a short wash in SSCT.

Incubation with antibody solutions is performed for 30 min each at 37 °C in a moist chamber. After every detection step, slides are washed three times in SSCT for 3 min each at room temperature. Agitate on a platform shaker, set to slow speed.

The following protocols show the incubation steps for detecting biotinylated and digoxigenin-labelled probes.

(a) Detection of biotin with FITC

- 1 avidin–FITC, diluted 1 : 500
- 2 biotinylated anti-avidin, 1 : 100
- 3 avidin–FITC, 1 : 500

Signals might be sufficiently strong (particularly from YACs and cosmids) with only one layer of avidin–FITC (i.e. omitting steps 3 and 4). Detection is then completed by three washes in SSCT and PBS as described below. If signals are too weak, they can then be amplified using steps 2 and 3 (see Protocol 49c).

(b) Detection of biotin with Texas red on chromosomes banded with anti-BrdU–FITC

This protocol is used for obtaining hybridization signals simultaneously with a replication G- or R-banding pattern on chromosomes in which BrdU was incorporated during late or early S-phase (see Section 9.2.1.)

- 1 avidin–Texas red, diluted 1 : 500
- 2 biotinylated anti-avidin, 1 : 100
- 3 avidin–Texas red, 1 : 500
- 4 anti-BrdU–FITC, 1 : 10 (use 50 µl under a 22×50 mm coverslip)

Note: Anti-BrdU–FITC can deteriorate within a few days unless stored frozen. As this antibody should not be frozen and thawed more than once, the stock solution should be frozen in small aliquots (5 µl) for single use. This antibody diluted in 0.9% NaCl, 0.2% Tween 20 results in a brighter banding pattern than is obtained with SSCTM as dilution buffer.

Detection of the probe in green with avidin–FITC on a red banding pattern obtained with mouse anti-BrdU and Texas red conjugated anti-mouse antibody results in brighter and more distinct probe signals but less clear banding pattern.

(c) Detection of biotin with Texas red and digoxigenin with FITC

- 1 mouse anti-digoxigenin (monoclonal), diluted 1 : 250, and avidin–Texas red, 1 : 500
- 2 FITC-labelled sheep anti-mouse, 1 : 50
- 3 biotinylated anti-avidin, 1 : 100
- 4 avidin–Texas red, 1 : 500

The last detection step in each protocol is followed by 1×5 min wash in SSCT and 2×5 min wash in PBS. Slides are then dehydrated in an

ethanol series (70%, 95%, absolute), air-dried and mounted in Citifluor containing DAPI (0.2 µg ml⁻¹) as counterstain. If no probe has been detected in red (e.g. Protocol 49a), propidium iodide (0.5 µg ml⁻¹) can be used instead, or in addition to, DAPI. Alternative mounting media are also available—for example, Vectashield from Vector Laboratories, a self-prepared mixture containing 22 mg 1,4-diazobicyclo (2.2.2.) octane (DABCO) in 1 ml 20 mM NaHCO₃ (pH 8.0), 75% glycerol [54], or 10 mg ml⁻¹ *p*-phenylenediamine in PBS mixed 1:9 with glycerol and adjusted to pH 8.0 with 0.5 M carbonate-bicarbonate buffer (pH 9.0) [55].

Very weak probe signals can be further amplified. Mounting medium and counterstain are first removed by rinsing the slide with methanol. The air-dried slide is then rehydrated in SSCT and incubated with SSCTM ('blocking' step). One round of signal amplification as in Protocols 49a and 49b is performed using biotinylated antiavidin (step 3) followed by step 4 (fluorochrome-conjugated avidin). More than two rounds of amplification is usually not beneficial since background staining will increase as well. In Protocol 49c, the signal of the biotinylated probe is amplified as in Protocol 49b. For amplification of the signal obtained with the digoxigenin-labelled probe, additional antibodies are necessary (e.g. FITC-labelled anti-sheep immunoglobulin).

.....

Troubleshooting

No signal visible

If other probes on different slides have worked in the same experiment, insufficient purified DNA or incorrect DNA concentrations are the most likely reasons.

- *Repurify the probe DNA before labelling.*
- *Check the DNA concentration for probe labelling.*
- *Check the labelling.*

For small DNA probes, additional signal amplification might be helpful. This is, however, only useful when the background is low.

- *Use more probe DNA for hybridization.*

High background staining of chromosomes and nuclei

This can be caused by insufficiently competed repetitive sequences. Increase the amount of Cot-1 DNA (up to 10-fold) for prehybridization.

A second approach is to increase the stringency of hybridization or posthybridization washes.

- *Prewarm the slide to 37 °C before applying the probe for hybridization to avoid annealing of the probe to nonspecific sequences until the temperature on the slide has reached the hybridization temperature.*

- *Increase the formamide concentration in the hybridization mixture (55% instead of 50%).*
- *For post-hybridization washes, use 1 ×SSC or even 0.1 ×SSC instead of 2 ×SSC in the second solution, which does not contain formamide.*

High background staining also between chromosomes and nuclei

This may be caused by hybridization of the probe to RNA. Pretreat the slides with RNase.

If the specific signals are bright enough do not use signal amplification, as this will increase the background staining.

The chromosomes are fuzzy and swollen

Poor morphology of chromosomes more likely reflects imperfect conditions during harvesting and slide making rather than non-optimal *in situ* hybridization, especially if overheating (>75 °C) during the denaturation of chromosomal DNA can be excluded.

Cytoplasm surrounding the chromosomes can take up moisture and prevent proper ageing. Beside background problems and insufficient hybridization results, cytoplasm therefore can also lead to swollen chromosomes. The speed of fixative evaporation, which is influenced by the air humidity during slide making, is a critical factor for the quality of metaphase spreads [56]. Cytoplasm usually can be reduced by preparing metaphase spreads in an area with increased humidity or by using fixative with a lower percentage of methanol.

- *Make the slides over a hot water bath or on a moist paper towel.*
- *Use fixative with a higher percentage of acetic acid (5 : 2 instead of 3 : 1).*
- *Slides may be rinsed with fixative shortly before complete evaporation of the first fixative.*

.....

References

1

Pardue, M.L. & Gall, J.G. (1969) Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proc. Natl Acad. Sci. USA* **64**, 600–604.

2

John, H., Birnstiel, M. & Jones, K. (1969) RNA–DNA hybrids at the cytological level. *Nature* **223**, 582–587.

3

Bauman, J.G.J., Wiegant, J., Borst, P. & van Duijn, P. (1980) A new method for fluorescence microscopical localization of specific DNA sequences by *in situ* hybridization of fluorochrome labelled RNA. *Exp. Cell Res.* **138**, 485–490.

4

Langer, P.R., Waldrop, A.A. & Ward, D.C. (1981) Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes. *Proc. Natl Acad. Sci. USA* **78**, 6633–6637.

5

van Prooijen-Knegt, A.C., van Hoek, J.F.M., Bauman, J.G.J. *et al.* (1982) *In situ* hybridization of DNA sequences in human metaphase chromosomes visualized by an indirect fluorescent immunocytochemical procedure. *Exp. Cell Res.* **141**, 397–407.

6

Landegent, J.E., Jansen in de Wal, N., Baan, R.A. *et al.* (1984) 2-Acetylaminofluorene-modified probes for the indirect hybridochemical detection of specific nucleic acid sequences. *Exp. Cell Res.* **153**, 61–72.

7

Tchen, P., Fuchs, R.P.P., Sage, E. & Leng, M. (1984) Chemically modified nucleic acids as immunodetectable probes in hybridization experiments. *Proc. Natl Acad. Sci. USA* **81**, 3466–3470.

8

Nederlof, P.M., Robinson, D., Abuknesha, R. *et al.* (1989) Three-color fluorescence *in situ* hybridization for the simultaneous detection of multiple nucleic acid sequences in interphase nuclei and chromosomes. *Cytometry* **10**, 20–27.

- 9 Nederlof, P.M., van der Flier, S., Wiegant, J. *et al.* (1990) Multiple fluorescence *in situ* hybridization. *Cytometry* **11**, 126–131.
- 10 Ried, T., Baldini, A., Rand, T.C. & Ward, D.C. (1992) Simultaneous visualization of seven different DNA probes by *in situ* hybridization using combinatorial fluorescence and digital imaging microscopy. *Proc. Natl Acad. Sci. USA* **89**, 1388–1392.
- 11 Wiegant, J., Wiesmeijer, C.C., Hoovers, J.M.N. *et al.* (1993) Multiple and sensitive fluorescence *in situ* hybridization with rhodamine-, fluorescein-, and coumarin-labeled DNAs. *Cytogenet. Cell Genet.* **63**, 73–76.
- 12 Lawrence, J.B., Villnave, C.A. & Singer, R.H. (1988) Sensitive, high-resolution chromatin and chromosome mapping *in situ*: presence and orientation of two closely integrated copies of EBV in a lymphoma line. *Cell* **52**, 51–61.
- 13 Lawrence, J.B., Singer, R.H. & McNeil, J.A. (1990) Interphase and metaphase resolution of different distances within the human dystrophin gene. *Science* **249**, 928–931.
- 14 Trask, B., Pinkel, D. & van den Engh, G. (1989) The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. *Genomics* **5**, 710–717.
- 15 Trask, B.J., Massa, H., Kenwrick, S. & Gitschier, J. (1991) Mapping of human chromosome Xq28 by two-color fluorescence *in situ* hybridization of DNA sequences to interphase cell nuclei. *Am. J. Hum. Genet.* **48**, 1–15.
- 16 Senger, G., Ragoussis, J., Trowsdale, J. & Sheer, D. (1993) Fine mapping of the MHC class II region within 6p21 and evaluation of probe ordering using interphase fluorescence *in situ* hybridization. *Cytogenet. Cell Genet.* **64**, 49–53.
- 17 Shipley, J.M., Jones, T.A., Patel, K. *et al.* (1993) Ordering of probes surrounding the Ewing's sarcoma breakpoint on chromosome 22 using fluorescent *in situ* hybridization to interphase nuclei. *Cytogenet. Cell Genet.* **64**, 233–239.
- 18 Wiegant, J., Kalle, W., Mullenders, L. *et al.* (1992) High-resolution *in situ* hybridization using DNA halo preparations. *Hum. Mol. Genet.* **1**, 587–591.
- 19 Heng, H.H.Q., Squire, J. & Tsui, L.-C. (1992) High-resolution mapping of mammalian genes by *in situ* hybridization to free chromatin. *Proc. Natl Acad. Sci. USA* **89**, 9509–9513.
- 20 Parra, I. & Windle, B. (1993) High-resolution visual mapping of stretched DNA by fluorescent hybridization. *Nature Genet.* **5**, 17–21.
- 21 Fidlerová, H., Senger, G., Kost, M. *et al.* (1994) Two simple procedures for releasing chromatin from routinely fixed cells for fluorescence *in situ* hybridization. *Cytogenet. Cell Genet.* **65**, 203–205.
- 22 Senger, G., Jones, T.A., Fidlerová, H. *et al.* (1994) Released chromatin: linearized DNA for high resolution fluorescence *in situ* hybridization. *Hum. Mol. Genet.* **3**, 1275–1280.
- 23 Ried, T., Lengauer, C., Cremer, T. *et al.* (1992) Specific metaphase and interphase detection of the breakpoint region in 8q24 of Burkitt lymphoma cells by triple-colour fluorescence *in situ* hybridization. *Genes Chrom. Cancer* **4**, 69–74.
- 24 Taylor, C., Patel, K., Jones, T. *et al.* (1993) Diagnosis of Ewing's sarcoma and peripheral neuroectodermal tumour based on the detection of t(11; 22) using fluorescence *in situ* hybridization. *Br. J. Cancer* **67**, 128–133.
- 25 Baldini, A., Ross, M., Nizetic, D. *et al.* (1992) Chromosomal assignment of human YAC clones by fluorescence *in situ* hybridization: use of single-yeast-colony PCR and multiple labeling. *Genomics* **14**, 181–184.
- 26 Garson, J.A., van den Berghe, J.A. & Kemshead, J.T. (1987) Novel non-isotopic *in situ* hybridization technique detects small (1 kb) unique sequences in routinely G-banded human chromosomes: fine mapping of N-myc and β -NGF genes. *Nucleic Acid Res.* **15**, 4761–4770.
- 27 Cherif, D., Bernard, O. & Berger, R. (1989) Detection of single-copy genes by nonisotopic *in situ* hybridization on human chromosomes. *Hum. Genet.* **81**, 358–362.
- 28 Lawrence, J.B. (1990) A fluorescence *in situ* hybridization approach for gene mapping and the study of nuclear organization. In *Genome Analysis*. Vol. 1: *Genetic and Physical Mapping* (Davies K.E. & Tilghman S., eds), (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).
- 29 Viegas-Péquignot, E., Berrard, S., Brice, A. *et al.* (1991) Localization of a 900-bp-long fragment of the human choline acetyltransferase gene to 10q11.2 by non-radioactive *in situ* hybridization. *Genomics* **9**, 210–212.
- 30 Mezzanotte, R., Vanni, R., Flore, O. *et al.* (1988) Ageing of fixed cytological preparations produces degradation for chromosomal DNA. *Cytogenet. Cell Genet.* **48**, 60–62.
- 31 Schweizer, D. (1976) Reverse fluorescent chromosome banding with chromomycin and DAPI. *Chromosoma* **58**, 307–324.
- 32 Vogel, W., Autenrieth, M. & Speit, G. (1986) Detection of bromodeoxyuridine-incorporation in mammalian chromosomes by a bromodeoxyuridine-antibody. I. Demonstration of replication patterns. *Hum. Genet.* **72**, 129–132.
- 33 Perry, P. & Wolff, S. (1974) New Giemsa method for the differential staining of sister chromatids. *Nature* **257**, 156–158.
- 34 Rønne, M. (1984) Fluorouracil synchronization of human lymphocyte cultures. Induction of high resolution R-banding by simultaneous *in vitro* exposure to 5-bromodeoxyuridine/Hoechst 33258. *Hereditas* **101**, 205–208.
- 35 Fan, Y.-S., Davis, L.M. & Shows, T.B. (1990) Mapping small DNA sequences by fluorescence *in situ* hybridization directly on banded metaphase chromosomes. *Proc. Natl Acad. Sci. USA* **87**, 6223–6227.
- 36 Baldini, A. & Ward, D.C. (1991) *In situ* hybridization banding of human chromosomes with Alu-PCR products: A simultaneous karyotype for gene mapping studies. *Genomics* **9**, 770–774.
- 37 Nelson, D.L., Ledbetter, S.A., Corbo, L. *et al.* (1989) Alu polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex

- DNA sources. *Proc. Natl Acad. Sci. USA* **86**, 6686–6690.
- 38 Willard, H.F. & Wayne, J.S. (1987) Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends Genet.* **3**, 192–198.
 - 39 Landegent, J.E., Jansen in de Wal, N., Dirks, R.W. *et al.* (1987) Use of whole cosmid cloned genomic sequences for chromosomal localization by non-radioactive *in situ* hybridization. *Hum. Genet.* **77**, 366–370.
 - 40 Lichter, P., Tang, C.C., Call, K. *et al.* (1990) High-resolution mapping of human chromosome 11 by *in situ* hybridization with cosmid clones. *Science* **247**, 64–69.
 - 41 Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
 - 42 Lengauer, C., Green, E.D. & Cremer, T. (1992) Fluorescence *in situ* hybridization of YAC clones after Alu-PCR amplification. *Genomics* **13**, 826–828.
 - 43 Feinberg, A.P. & Vogelstein, B. (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**, 266–267.
 - 44 Lo, Y.M.D., Mehal, W.Z. & Fleming, K.A. (1988) Rapid production of vector-free biotinylated probes using the polymerase chain reaction. *Nucleic Acids Res.* **16**, 8719.
 - 45 Sverdlov, E.D., Monastyrskaya, G.S., Guskova, L.I. *et al.* (1974) Modification of cytidine residues with a bisulfite-O-methylhydroxylamine mixture. *Biochim. Biophys. Acta* **340**, 153–165.
 - 46 Bauman, J.G.J., Wiegant, J. & van Duijn, P. (1983) The development, using poly (Hg-U) in a model system, of a new method to visualize cytochemical hybridization in fluorescence microscopy. *J. Histochem. Cytochem.* **31**, 571–578.
 - 47 Hopman, A.H.N., Wiegant, J. & van Duijn, P. (1985) A new hybridocytochemical method based on mercurated nucleic acid probes and sulfhydryl-hapten ligands. II. Effects of variations on the *in situ* detection of mercurated probes. *Histochemistry* **84**, 179–185.
 - 48 Hulspar, R. & Bauman, J.G.J. (1992) The use of fluorescent *in situ* hybridization for the analysis of nuclear architecture by confocal microscopy. *Cell Biol. Int. Repts* **16**, 739–747.
 - 49 Moyzis, R.K., Buckingham, J.M., Cram, L.S. *et al.* (1988) A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proc. Natl Acad. Sci. USA* **85**, 6622–6626.
 - 50 Shiloh, Y. & Cohen, M.M. (1978) An improved technique of preparing bone marrow specimens for cytogenetic analysis. *In Vitro* **14**, 510–515.
 - 51 Vogel, W., Schempp, W. & Sigwarth, I. (1978) Comparison of thymidine, fluorodeoxyuridine, hydroxyurea, and methotrexate blocking at the G1/S phase transition of the cell cycle, studied by replication patterns. *Hum. Genet.* **45**, 193–198.
 - 52 Wiley, J.E., Sargent, L.M., Inhorn, S.L. & Meisner, L.F. (1984) Comparison of prometaphase chromosome techniques with emphasis on the role of colcemid. *In Vitro* **20**, 937–941.
 - 53 Raap, A.K., Marijnen, J.G.J., Vrolijk, J. & van der Ploeg, M. (1986) Denaturation, renaturation, and loss of DNA during *in situ* hybridization procedures. *Cytometry* **7**, 235–242.
 - 54 Scherthan, H., Köhler, M., Vogt, P. *et al.* (1992) Chromosomal *in situ* hybridization with double-labeled DNA: signal amplification at the probe level. *Cytogenet. Cell Genet.* **60**, 4–7.
 - 55 Johnson, G.D. & Nogueira Araujo, G.M.C. (1981) A simple method of reducing the fading of immunofluorescence during microscopy. *J. Immunol. Meth.* **43**, 349–350.
 - 56 Lawce, H.J. & Brown, M.G. (1991) Harvesting, slide-making, and chromosome elongation techniques. In *The ACT Cytogenetics Laboratory Manual* (Barch M.J., ed.), (Raven Press, New York), pp. 31–105.
 - 57 Fan, Y.S., Sasi, R., Lee, C. *et al.* (1992) Mapping of 50 cosmid clones isolated from a flow-sorted human X chromosome library by fluorescence *in situ* hybridization. *Genomics* **14**, 542–545.
 - 58 Trask, B., Fertitta, A., Christensen, M. *et al.* (1993) Fluorescence *in situ* hybridization mapping of human chromosome 19: cytogenetic band location of 540 cosmids and 70 genes or DNA markers. *Genomics* **15**, 133–145.
 - 59 Selig, S., Okumura, K., Ward, D.C. & Cedar, H. (1992) Delineation of DNA replication time zones by fluorescence *in situ* hybridization. *EMBO J.* **11**, 1217–1225.
 - 60 Kitsberg, D., Selig, S., Brandeis, M. *et al.* (1993) Allele-specific replication timing of imprinted gene regions. *Nature* **364**, 459–463.
 - 61 Bischoff, A., Albers, J., Kharboush, I. *et al.* (1993) Differences of sizes and shape of active and inactive X-chromosome domains in human amniotic fluid cell nuclei. *Microsc. Res. Techn.* **25**, 68–77.
 - 62 Boyle, A.L., Lichter, P. & Ward, D.C. (1990) Rapid analysis of mouse-hamster hybrid cell lines by *in situ* hybridization. *Genomics* **7**, 127–130.
 - 63 Villnave Johnson, C., Singer, R.H. & Lawrence, J.B. (1991) Fluorescent detection of nuclear RNA and DNA: Implications for genome organization. *Meth. Cell Biol.* **35**, 73–99.
 - 64 Jauch, A., Daumer, C., Lichter, P. *et al.* (1990) Chromosomal *in situ* suppression hybridization of human gonosomes and autosomes and its use in clinical cytogenetics. *Hum. Genet.* **85**, 145–150.
 - 65 Taylor, C.P., McGuckin, A.G., Bown, N.P. *et al.* (1994) Rapid detection of prognostic genetic factors in neuroblastoma using fluorescence *in situ* hybridization on tumour imprints and bone marrow smears. *Br. J. Cancer* **69**, 445–451.
 - 66 Graf, D., Korthäuer, U., Mages, H.W. *et al.* (1992) Cloning of TRAP, a ligand for CD40 on human T cells. *Eur. J. Immunol.* **22**, 3191–3194.
 - 67 Korthäuer, U., Graf, D., Mages, H.W. *et al.* (1993) Defective expression of T-cell CD40 ligand causes X-linked immunodeficiency with hyper-IgM. *Nature* **361**, 539–541.
 - 68 Kroccek, R.A., Graf, D., Brugnani, D. *et al.* (1994) Defective expression of CD40 ligand on T cells causes X-linked immunodeficiency with hyper-IgM (HIGM1). *Immunol. Rev.* **138**, 39–59.
 - 69 Padayachee, M., Feighery, C., Finn, A. *et al.* (1992) Mapping of the X-linked form of hyper-IgM syndrome (HIGM1) to Xq26 by close linkage to HPRT. *Genomics* **14**, 551–553.

Chapter 10 Chromosome painting

Lyndal Kearney

MRC Molecular Haematology Department, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK

10.1	Introduction, 242
10.2	Resources available, 243
10.2.1	Chromosome libraries, 243
10.2.2	Interspersed repetitive sequence polymerase chain reaction, 244
10.2.2.1	Alu-PCR, 244
10.2.3	DOP-PCR, 245
10.2.4	Flow sorting of abnormal chromosomes, 245
10.2.5	Microdissection and FISH, 246
10.2.6	Commercial painting probes, 246
10.2.7	Multicolour painting, 246
10.3	Competitive in situ suppression hybridization, 247
10.4	Applications, 247
10.4.1	Detection of chromosomal abnormalities, 247
10.4.2	Interphase cytogenetics, 248
10.5	Discussion, 248
10.5.1	Limitations of chromosome painting probes, 248
10.5.2	Future prospects, 249
	Protocol 50
	Nick translation, 250

Protocol 51
Alu-PCR amplification of flow-sorted chromosomes, 252
Protocol 52
DOP-PCR amplification of flow-sorted normal and abnormal chromosomes, 253
Protocol 53
Competitive (or chromosomal) in situ suppression hybridization, 255
Protocol 54
Detection of hybridized labelled probe, 257
Troubleshooting, 259
Amplification of negative control after first round DOP-PCR, 259
Cell lost from slide, 259
Poor chromosome morphology, 259
No hybridization signal, 259
High background, 259
References, 260

10.1 Introduction

Routine cytogenetic analysis has undergone a series of improvements over the past 40 years, from the establishment of banding techniques to the development of prometaphase chromosome preparations for high-resolution banding analysis. Such high-resolution analysis can detect small chromosomal deletions of 1–2 Mb, and has been important in the detection of microdeletion syndromes such as Prader–Willi, Miller–Dieker and α -thalassaemia mental retardation syndromes. Fluorescence *in situ* hybridization (FISH) with specific disease-region probes can detect chromosome abnormalities not visible by high-resolution cytogenetic analysis and advances in FISH technology have revolutionized gene mapping and genome analysis in a number of ways (see Chapter 9). However, in a large number of genetic disorders and in many forms of cancer, the specific defect is not known. Chromosome painting is one application of FISH that helps to alleviate this problem, improving the accuracy of cytogenetic studies and closing the gap between cytogenetic and molecular analysis.

A set of DNA probes derived from a single chromosome type can be used to delineate the whole or part of that chromosome by *in situ* hybridization to chromosomal DNA. The direct visualization of specific chromosomes by fluorescent detection of hybridized labelled whole-chromosome probes has led to the term ‘chromosome painting’ and to the whole-chromosome specific probes being called ‘paints’ [1–4]. The advent of competitive *in situ*

suppression hybridization (CISSH) (also known as chromosomal *in situ* suppression hybridization) allows the removal of ubiquitous repeat sequences within the whole-chromosome probes before their use as chromosome paints.

At the simplest level, total human DNA, biotin-labelled and used as a probe to human–hamster hybrid cell lines can be used to paint the human complement of these cell lines [5]. Chromosomes purified by fluorescence-activated flow sorting (FACS) analysis (see Chapter 12) have been used for the construction of whole-chromosome specific libraries [6,7]. Whole-chromosome painting probes prepared from these libraries consist of many different clones distributed more or less evenly over the chromosome [8], and chromosome paints derived in this way are now commercially available (Table 10.1).

Another approach has been to apply the technique of interspersed repetitive sequence polymerase chain reaction (IRS-PCR) or degenerate oligonucleotide PCR (DOP-PCR) (see also Chapter 11, Protocol 63) to monochromosomal somatic cell hybrid DNA (Chapter 14) or flow-sorted chromosome fractions in order to amplify chromosome-specific sequences for use as painting probes. A variation on the chromosome painting theme (so-called reverse painting) can be used to identify *de novo* unbalanced chromosomal rearrangements by flow-sorting the abnormal chromosomes and using the labelled abnormal chromosome as a probe to normal metaphase chromosomes [9–12].

An innovation combining FISH and chromosome microdissection (Chapter 11) makes it theoretically possible to obtain region-specific paints for any part of the human genome. The microFISH technique [13] employs chromosome microdissection and amplification by PCR (Chapter 11, Protocols 62 and 63) to provide a probe that can be used as a band-specific paint when used in CISSH experiments. The ability to target chromosome regions by microdissection and enzymatic amplification provides a new way of determining the origin of otherwise unidentifiable chromosome segments [14–16].

FISH with whole-chromosome painting probes allows the accurate identification of structural chromosome aberrations in metaphase cells with complex karyotypes (i.e. with multiple structural rearrangements) such as are frequently found in solid tumours and cell lines, and from haematological malignancies, which are not always amenable to conventional cytogenetic analysis by Giemsa (G)-banding (Chapter 7).

Combinatorial and ratio labelling of probes has increased the number of target sequences that can

Chromosome painting can be used to:

- identify human chromosomes in somatic cell hybrids
- identify specific chromosomes in metaphase and interphase cells
- accurately identify structural chromosomal rearrangements in metaphase cells with complex karyotypes, such as tumour cells and leukaemic cells
- distinguish translocations and other chromosomal rearrangements that cannot be detected by conventional cytogenetic methods such as G-banding
- determine the origin of otherwise unidentifiable chromosomal fragments
- detect abnormalities of chromosomal regions for which no locus-specific probes are available
- screen rapidly for damage to chromosomes caused by ionizing radiation
- identify *de novo* chromosomal abnormalities
- identify cross-species regions of homology

Applications box 10.1

Table 10.1 Commercially available probes for FISH.

Appligene Oncor (biotin and digoxigenin-labelled probes): <i>Centromere-specific (alpha and beta satellite) probes for all human chromosomes</i> <i>A range of chromosome-specific telomere probes</i> <i>Coatosome whole-chromosome paints for all human chromosomes</i> <i>Probes for microdeletion syndromes including:</i> Miller–Dieker (17p13.3), Cri-du-chat (5p15), Wolf–Hirschorn (4p16.3), DiGeorge (22q11.2), Prader–Willi / Angelman (15q11–q13), Charcot–Marie–Tooth (17p11.2), Smith–Magenis (17p11.2), Williams (Elastin gene)(7q11.23) <i>Region-specific unique sequence probes including:</i> 5q31, 6q27, 8q21, 10q22, 13q14, 19q13, 21q22, Xq13.2 <i>Probes for oncology/leukaemia research including:</i> p53 (17p13.3), retinoblastoma (13q14), N-myc (2p23–p24), HER-2 / neu (17q11.2–q12), Mbcrl (22q11.2), abl (9q34), MLL (11q23) <i>Two-colour translocation detection probes:</i> Mbcrl/abl (major breakpoint), t(15;17), mbcrl/abl (minor breakpoint), iso(17q) Vysis Ltd (probes directly labelled with Spectrum Orange, Spectrum Green, Spectrum Aqua): Whole-chromosome paints for all human chromosomes Centromere-specific probes (chromosome enumerator probes) for all human chromosomes FISH probes for aneuploidy detection in prenatal and postnatal genetics Microdeletion syndrome probes (as above) Oncology / leukaemia probes (as above, including two-colour translocation probes) CGH reagents Cambio Ltd (biotin, FITC, Cy3, and Cy5 labelled paints): DOP-PCR derived whole-chromosome paints (all human chromosomes) Mouse whole-chromosome probes (1–12, 14–19, X and Y) Centromere-specific (alpha and beta satellite) probes for all human chromosomes Cytocell Ltd Chromoprobe Multiprobe system for the simultaneous analysis of all human chromosomes whole-chromosome paints chromosome-specific centromere probes
--

be detected in a single multicolour hybridization experiment [17–20]. In the combinatorial approach, probes are labelled with varying ratios of different haptens (e.g. biotin digoxigenin) and their hybridization detected using combinations of fluorescent polyclonal and monoclonal antibodies [19,20]. Ratio labelling uses different ratios of differently labelled probes and has been used to paint half of the human chromosome complement in 12 different colours [21]. The combinatorial labelling approach has now been refined to discriminate 27 different colours [22,23]. This very significant advance was achieved by direct labelling of whole chromosome paints (and some chromosome arm paints) with a combination of five different fluorochromes (in addition to the DAPI counterstain) and detection using specific, narrow band-pass filter sets and computer software to discriminate the spectral signature for each chromosome paint.

The ability to visualize multiple colours in multiplex hybridizations with paints and YACs has led to the concept of ‘chromosomal bar codes’,

specific patterns of differentially labelled chromosomes, with the aim of constructing sets of probes tailored to specific diagnostic problems [24]. A combination of chromosome paints and YACs or cosmids can also be used as an alternative to reverse painting, where marker chromosomes cannot easily be separated by flow sorting, for example in leukaemic bone marrow metaphases [25].

10.2 Resources available

10.2.1 Chromosome libraries

Whole-chromosome painting probes have been prepared from chromosomes enriched for a single type by FACS analysis (Chapter 12). The purified flow-sorted chromosome fractions are digested with restriction enzymes and cloned into a bacterial vector. DNA extracted from the pooled chromosome-specific library probes is labelled with a hapten or fluorochrome to generate a complex probe suitable for FISH. It is desirable for all gene mapping

applications, including chromosome painting, that human chromosome-specific libraries should be representative of the original chromosome in complexity.

Earlier libraries were produced by cloning flow-sorted chromosomal DNA into phage vectors, limiting the size of insert cloned. Libraries of larger fragments are now available cloned in cosmid vectors [6]. The flow-sorted chromosome-specific libraries cloned in Charon 21 phage vectors were not suitable for use as chromosome painting probes because of the high ratio of vector to insert (approx. 90% of the DNA in the library is from the vector). This problem has now been largely overcome by *HindIII* digestion of the original phage libraries and subcloning into plasmids [7].

These libraries are available for all human chromosomes, although their usefulness as chromosome painting probes is variable: the plasmid libraries for chromosomes 1, 4, 9, 11, 16, 18 and 20 do not hybridize to centromeres of their target chromosomes, and those for chromosomes 13, 14, 15, 21 and 22 cross-hybridize with the centromeres of all acrocentric chromosomes. Another approach allows the production of complex chromosome-specific libraries by linker-adaptor PCR [26]. DNA from flow-sorted chromosome fractions is digested with a frequently cutting restriction enzyme and ligated at each end to an adaptor oligonucleotide. This allows the fragments to be amplified using a primer for the adaptor sequence. Linker-adaptor libraries are now available for all human chromosomes [26].

The advantage of the plasmid libraries is that they can be grown up in bulk without any loss of complexity. However, purification and labelling for use as chromosome-painting probes is quite time consuming. Amplification and labelling of the PCR libraries is rapid and simple [26], but repeated amplification may lead to loss of complexity, and preparation of large quantities of labelled probe is expensive. Plasmid DNA from whole chromosome libraries is best purified by CsCl gradient centrifugation, or using Qiagen columns (Hybaid) or similar, followed by nick translation labelling as described in Protocol 50. The PCR libraries are labelled in a second round of PCR amplification [26] and purified through Sephadex G50 columns as in Protocol 50. The purified labelled probe is then ready for use in CISH procedures carried out as in Protocol 53. The amount of probe and competitor DNA depends on the vector-to-insert ratio of the library and the size of the chromosome, and may need to be determined empirically. The probe concentrations are usually in the range 100–500 ng (500 ng for A group chromosomes, 100–200 ng for E,

F and G group chromosomes), with 1–5 µg Cot-1 DNA, or 1–26 µg total human DNA. Detection of labelled probe is carried out as described in Protocol 54.

10.2.2 Interspersed repetitive sequence polymerase chain reaction

Large numbers of somatic cell hybrids are available as a resource in human gene mapping. These contain a single human (or translocation derivative) chromosome in a rodent background. Regular cytogenetic analysis of hybrids is essential as the human chromosomes are prone to deletion and rearrangement during culture; chromosome painting provides an accurate way of doing this. Total genomic DNA from somatic cell hybrids can be used as a paint to identify the human DNA present [4] and total genomic DNA from cell hybrids has also been used as a probe to confirm the presence of an isochromosome 12p in testicular germ-cell tumours [27]. However, the use of total cellular DNA as a probe has limitations, as only a small fraction represents the region of interest and the consequent sensitivity when used in FISH may be too low to detect minor chromosome segments. The technique of interspersed repetitive sequence polymerase chain reaction (IRS-PCR) was devised to amplify only the human sequences present in somatic cell hybrids, as a way of accurately characterizing the human content. This technique uses as primers oligonucleotides complementary to the human-specific consensus sequence of commonly occurring repetitive sequences such as Alu [28].

Alu repeats are short interspersed repeat sequences which are present in $\approx 10^6$ copies in the human genome [29]. The human Alu sequences are 280 bp long and although there is considerable variation, a consensus sequence has been established. The average spacing between Alu repeats is 4 kb, with Alu-rich regions having interAlu distances of 1 kb, and Alu-poor regions a distance of roughly 10 kb [30]. The use of a single Alu primer allows the amplification of the DNA sequences between two inverted Alu repeats, provided that these blocks of repeats are within a distance that can be bridged by PCR. Several groups have used IRS-PCR to characterize human chromosomes in interspecies hybrids, using either primers for human-specific Alu or for the L1 element of long interspersed repeats [31–33], and to isolate chromosome-specific probes [34].

10.2.2.1 Alu-PCR

The Alu-PCR technique can also be applied to small

numbers of flow-sorted chromosomes [35,37] as an aid to gene mapping. The purification of chromosomes by FACS analysis is described in Chapter 12. A similar methodology may be applied to the production of chromosome paints. We have produced chromosome painting probes for chromosomes 1–8, 17, 18, 19, 21 and 22 by amplification of small numbers (200–500) of flow-sorted chromosomes using a primer for the human Alu consensus sequence (see Plate 4). The direct amplification of small numbers of flow-sorted chromosomes has obvious advantages over the use of whole chromosome libraries for obtaining chromosome-painting probes. The sorting required to generate chromosome-specific DNA libraries is of the order of 1–2 weeks and carries with it the chance of introducing sequence contamination as a result of chromosome damage. In comparison, it only takes minutes to sort sufficient chromosomes for the generation of Alu-PCR probes, thereby reducing the risk of chromosome fragmentation.

The chromosome-specific paints obtained using Alu primers generate reproducible reverse (R)-banding patterns when hybridized to metaphase chromosomes [37]. These patterns reflect the relative richness of Alu sequences in G-negative bands [38]. When cloned region-specific probes are applied at the same time, the R-banding pattern can be used to assign the cloned probes to specific chromosomal bands. The most obvious difference between Alu-PCR banding and conventional R-banding is the lack of any staining in the regions of constitutive heterochromatin (usually at the centromeres and the long arm of the Y) when using Alu-PCR products.

10.2.3 DOP-PCR

Another method for obtaining whole chromosome paints by PCR of flow-sorted chromosome fractions involves the random amplification of DNA at many sites in the genome using a partially degenerate oligonucleotide primer [9]. The 3' end of the primer has six specified bases allowing amplification at frequently occurring sites at the low annealing temperature in the first round of amplification (Fig.10.1, see also Fig.11.4 in Chapter 11). The presence of six degenerate oligonucleotides 5' to the specified sequence allows a more general amplification than would occur with a non-degenerate primer. The 5' end has another six specified bases which anneal to previously amplified sequences in later PCR cycles which can be carried out at higher temperatures. As this amplification does not rely on the orientation of repeat sequences, chromosome-specific paints derived by DOP-PCR give a more

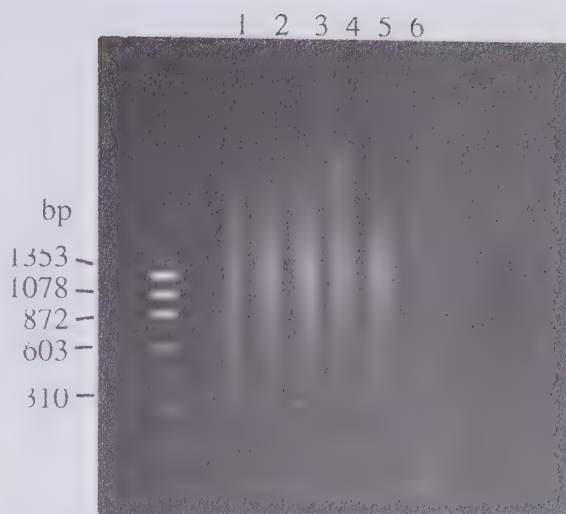


Fig. 10.1 First round DOP-PCR amplification of flow-sorted chromosomes. The amplified chromosomes are in lanes 1–5; lane 6 is the negative control. The PCR reaction was carried out as described in Protocol 52. Ten microlitres of product were run on a 1.2% agarose gel (containing ethidium bromide) at 50 V for 1 h. The marker is ϕ X174 *Hae*III. Amplified products in lanes 1–5 range in size from 300 bp to > 1353 bp. Products in this size range are commonly seen after second round amplification and labelling, and are cut to size with 5 μ M DNase 1 for a further 30 min to 1 h. There is no amplification in the negative control lane (lane 6).

even coverage of chromosomes than Alu-PCR derived paints. This technique has the added advantage of not being species specific and can be used to amplify DNA from any source.

10.2.4 Flow sorting of abnormal chromosomes

The application of chromosome-specific paints to complex karyotypes can be laborious because of the numerous combinations of chromosome paints required. An alternative approach to the identification of an abnormal chromosome has been to isolate the abnormal chromosome by flow sorting, followed by DOP-PCR amplification of sequences from the abnormal chromosome and hybridization of these probes to normal chromosomes. This technique has been called 'reverse chromosome painting' [11,12]. Provided that a cell line containing the abnormality is available to provide a source of sufficient chromosomes, and that the unidentified abnormal chromosome differs sufficiently in DNA content and sequence from its normal counterpart, it can be separated by flow cytometry. Alu-PCR of flow-sorted abnormal chromosomes has been used to derive paints from the derivative chromosome

9 from a cell line carrying the Philadelphia translocation $t(9;22)(q34;q11)$, from the derivative chromosome 11 of a constitutional $t(11;22)(q23;q11)$ and from a chromosome 11 with a partial deletion of the long arm [39]. Reverse chromosome painting also provides a way of deriving chromosome region- or band-specific paints. Examples of region-specific paints derived in this way are shown in Plate 5. A modification of reverse painting has used tumour-cell genomic DNA as a probe for CISSH [40]. This tumour-derived chromosome paint identified the chromosomal location of DNA sequences that were amplified in the tumour cell.

10.2.5 Microdissection and FISH

The identification of unknown chromosome segments in complex karyotypes requires 24 different chromosome-specific painting probes (one for each pair of autosomes and one for each sex chromosome). Although multicolour FISH analysis has now advanced to allow the simultaneous detection of 27 different colours [22,23], some structural rearrangements (for example, some pericentric inversions) are not visualized using whole-chromosome paints. In addition, whole-chromosome paints do not give any regional information, for example about which genes may be deleted or duplicated.

Detection of some disease-specific translocations is possible with single-copy probes for well-characterized loci (see Table 10.1). However, for the majority of chromosome rearrangements, specific probes are not available. A recent innovation now makes it possible in principle to obtain region-specific probes for any part of the human genome, by microdissection of human chromosomes with direct enzymatic amplification of the microdissected DNA fragments [13]. Approximately 25–30 chromosome fragments are microdissected from the region of interest and the DNA amplified by DOP-PCR [9] (see Chapter 11, Protocols 62 and 63). The purified PCR products are then labelled with biotin for hybridization to normal human metaphase cells. This technique has been used to identify a translocation and deletion chromosome in a malignant melanoma cell line [13] and to create a band-specific library for 6q21, a region frequently deleted in malignant melanoma [41]. Microdissection and FISH has also been used to show that terminal deletions of 6q are telomeric translocations [42] and to detect variant Philadelphia translocations in chronic myeloid leukaemia (CML) [43]. This strategy has great potential for the characterization of unidentifiable marker chromosomes such as

double minutes and homogeneously staining regions and of *de novo* constitutional chromosome abnormalities [14–16]. Microdissection has now been used to generate paints for all human chromosome arms (excluding the short arms of the acrocentric chromosomes) [44]. This makes it possible to identify pericentric inversions that may not be detected using whole chromosome paints. Chromosome microdissection for cloning and the production of region-specific paints is discussed in Chapter 11.

10.2.6 Commercial painting probes

An increasingly wide range of whole chromosome painting probes is now commercially available, ready labelled with biotin or digoxigenin or directly conjugated to fluorochromes (Table 10.1). Cambio supply relatively inexpensive DOP-PCR-derived paints labelled with biotin or FITC for chromosomes 1–9, 11–22, X and Y, as well as a range of detection kits. Vysis (formerly Imagenetics) supply directly fluorochrome-labelled whole-chromosome painting probes (derived from the pBS libraries) which work well and avoid the time-consuming immunochemical detection steps, making dual colour hybridization exceedingly simple (see Appendix III for addresses). Although these are sold as being compatible with specific Zeiss fluorescence filter sets, we have found that Nikon Optiphot fluorescence filter sets for rhodamine and FITC, as well as the single and dual channel filter blocks on the MRC 600 confocal microscope are suitable for BRL Spectrum Orange and Spectrum Green whole-chromosome paints (see Appendix IV, Table IV.4 for filter blocks). An ingenious device from Cytocell Ltd (the Chromoprobe Multiprobe system) provides an alternative to the multicolour painting approach. They supply a coverslip device with 24 different, fluorochrome-labelled painting probes already applied, and a gridded microscope slide for test metaphases. This allows the simultaneous detection of 24 different probes in a single colour.

The probe concentration and amount of competitor required for the different types of chromosome painting probe are given in Table 10.2.

10.2.7 Multicolour painting

Chromosome painting is particularly suited to simultaneous analysis of several targets. Simultaneous detection of two targets is achieved by labelling one probe with digoxigenin and one with biotin and carrying out a dual colour detection as in

Table 10.2 Probe and competitor concentrations for chromosome painting procedures.

Type of probe	Probe (ng μl^{-1})	Cot-1 DNA ($\mu\text{g } \mu\text{l}^{-1}$)
Chromosome library	10–50	0.1–0.5
Alu-PCR flow-sorted chromosomes	40	0.5
DOP-PCR flow-sorted chromosomes	10	0.625
Micro-FISH paint	10	0.1

Protocol 54. The combinatorial labelling method has recently been used for the simultaneous detection of 27 different targets, using probes labelled with varying ratios of five different fluorochromes [22]. However, at present the filter sets and software for this type of analysis are not widely available, and the simultaneous detection of two or three targets is a more realistic goal for most laboratories. A method for three-colour FISH is given in Chapter 9. Labelling ratio schemes for the simultaneous detection of three and five colours are given in Tables 10.3 and 10.4. Dual-colour detection is carried out as in Protocol 54. The availability of directly fluorochrome-conjugated nucleotides has simplified multicolour painting even further. Simultaneous detection of three targets is possible using one paint directly labelled with FITC (green), a second labelled with TRITC (red) and the third with a 1:1 mixture of FITC and TRITC (orange).

10.3 Competitive in situ suppression hybridization

Whole-chromosome painting probes require an additional step before hybridization in order to remove ubiquitous repetitive sequences. This is achieved by a short incubation, prior to hybridization, with unlabelled human competitor DNA, in the form of either total human DNA (placental DNA, sheared and sonicated to 50–300bp) or human Cot-1 DNA (Gibco-BRL). Cot-1 DNA is suitable for most purposes, but there may be occasions when moderately repeated sequences are

Table 10.3 Mixing ratio for simultaneous three-colour detection.

Probe	Biotin-Texas red	Digoxigenin-FITC
A	1	
B	1	1
C		1

not blocked by Cot-1. In these cases, total human DNA may be more suitable. It may be necessary to titrate the amount of total human DNA to determine the correct amount of competitor.

10.4 Applications

10.4.1 Detection of chromosomal abnormalities

The striking visualization of chromosome abnormalities using chromosome paints, as well as the ease and speed of the technique, makes chromosome painting an invaluable addition to the more traditional cytogenetic techniques (see Chapters 7 and 8). Chromosome painting readily detects translocations and numerical abnormalities in metaphase cells. The technique is particularly useful in cases where high quality G-banded analysis can be difficult, as in tumour cells, and for rapid screening of chromosome damage due to ionizing radiation [45]. Reverse chromosome painting allows the identification of *de novo* constitutional chromosome abnormalities [12], as well as the identification of amplified sequences, using whole tumour DNA as a probe [40]. Region-specific paints, obtained by microdissection or amplification of flow-sorted abnormal chromosomes, provide a novel way of detecting abnormalities of regions for which no locus-specific probes are available. In addition, microdissection and FISH provide a way of

Table 10.4 Mixing ratio for simultaneous five-colour detection.

Probe	Biotin-Texas red	Digoxigenin-FITC
A	1	
B	4	1
C	1	1
D	1	4
E		1

identifying abnormalities not amenable to other types of analysis, such as double minute chromosomes and homogeneously staining regions [16]. Multicolour FISH techniques have increased the number of targets that can be visualized simultaneously, thereby decreasing the number of procedures required to identify multiple abnormalities. This has applications in the detection of aneuploidy in metaphase cells both pre and postnatally, as well as in the characterization of complex tumour karyotypes and in assessing radiation damage.

10.4.2 Interphase cytogenetics

Individual chromosomes occupy discrete, relatively compact domains within interphase nuclei [46,47]. This has allowed the application of FISH to interphase cells, enabling a karyotype to be determined without the need for dividing cells, hence the term 'interphase cytogenetics'. However, Kuo *et al.* [48] reported a relatively low detection rate for trisomies in amniotic cells using chromosome paints. In a similar study, we found that only 30–50% of nuclei from the bone marrow of a patient with acute myeloid leukaemia (AML) exhibited three signals using a chromosome 8 painting probe. This was in contrast to 80% of bone marrow metaphase cells from the same patient showing three copies of chromosome 8 using the same painting probe (see Plate 6).

The low detection rate in interphase nuclei can be ascribed to the chromosomal orientation in the two-dimensional view of the nucleus under the microscope, or to an overlap of extended chromosome domains detected by painting probes. For this reason whole-chromosome paints are not suitable for the detection of numerical chromosomal abnormalities in interphase cells. The assessment of aneuploidy in interphase nuclei is best carried out using chromosome-specific repetitive probes such as alphoid centromere probes, or a pool of cosmid probes from the region of interest (e.g. the Down's syndrome critical region). These produce strong, tightly localized hybridization signals which allow rapid and accurate enumeration.

Structural chromosomal abnormalities in interphase nuclei have been detected with whole chromosome probes using FISH [1–3]. However, the problems of interpretation due to the overlap of extended chromosome domains still apply. Pinkel *et al.* [1] found that the three chromosome 4 signals corresponding to a t(4;11) were detected in only 50% of cells from a cell line carrying this translocation.

Separation of the signals representing whole chromosomes (e.g. due to lack of centromere sequences) also results in a high percentage of false positives using chromosome paints. Detection of specific chromosome translocations in interphase cells is more accurately achieved using locus-specific probes [49,50] (see Table 10.1).

It is now possible to combine cytogenetic and immunophenotypic information using interphase FISH and simultaneous fluorescent detection of cell-surface markers [51,52]. The ability to correlate chromosome abnormalities with the cell-surface antigens expressed in cells from a particular tumour has implications in determining the cell type involved in the tumour, as well as in monitoring response to therapy [53–55]. Although this technique has great potential, the problems of using chromosome paints for interphase cell analysis still apply, and chromosome-specific repetitive probes are therefore more suitable.

10.5 Discussion

10.5.1 Limitations of chromosome painting probes

The striking appearance of the 'painted' chromosomes and the simplicity and rapidity of the technique means that the use of chromosome painting probes is a valuable means of identifying numerical and structural chromosome abnormalities. However, their use is not applicable to every situation, and the precise specificity of each probe must be ascertained before use. The limitations in sensitivity of chromosome painting probes are not known and depend on the type of probe used. The pBS whole-chromosome libraries hybridize with varying degrees of intensity and specificity [7]. For example, the pBS-13, -14, -15, -21 and -22 libraries cross-hybridize with the centromeric regions of all chromosomes of this group, making metaphase analysis confusing and interphase analysis impossible. In addition, some of the libraries do not hybridize evenly along the length of the chromosome. The chromosome 1 paint from the pBS-1 library detects the 1p32-pter region only poorly, and translocations involving this region may be difficult to identify. The chromosome 9 paint from the pBS-9 library does not contain the 9q34-qter region, and therefore cannot be used to detect the Philadelphia chromosome. These deficiencies also apply to commercially available paints derived from these libraries. In general, the PCR libraries [26] give a more even intense staining, but also show cross-

hybridization of centromeric regions for the PCR-13, -14, -15, -21 and -22 libraries.

Chromosome paints derived from many of these libraries (both pBS and PCR) do not stain centromeres, causing problems in interpreting the number of signals in interphase nuclei and in the identification of small centric fragments. Similarly, Alu-PCR-derived paints do not cover the centromeric regions, because of a deficiency of Alu repeats in these regions. Chromosomes painted with Alu-PCR paints have an R-banded appearance, which can be advantageous in identifying the painted chromosome, but Alu-PCR paints may therefore not detect translocations involving the G-band regions. However, a modification of Alu-PCR amplification using two primers may be more sensitive and has recently been shown to detect a 1 Mb segment of 19p13 from a monochromosomal hybrid [56].

10.5.2 Future prospects

One of the most significant advances in chromosome painting technology has been reverse painting. This allows the identification of *de novo* chromosome rearrangements as well as the identification of marker chromosomes, provided that the abnormal chromosome is available in a cell line and is resolvable by flow sorting (for production of the probes). These requirements limit the use of reverse painting in the characterization of primary tumour material such as leukaemic bone marrow, as insufficient metaphase cells are available for flow sorting. An approach to this problem that does not require metaphase cell preparation is provided by the technique of comparative genomic hybridization [57] (see Chapter 8).

Advances in the production of region-specific paints by microdissection [14–16] and Alu-PCR amplified YAC sequences [25], as well as the introduction of multicolour labelling and detection protocols [21] and directly fluorochrome-labelled probes [58], mean that the accurate identification of complex chromosome abnormalities is now simplified. The developments of multifluor FISH and spectral karyotyping are certain to revolutionize the analysis of complex karyotypes, and provide insights into cross-species karyotyping [22,23,59]. The vivid multicolour images produced by this

rapidly developing technology (see Plates 4, 5 and 6) will ensure that chromosome painting continues to contribute to both scientific research and clinical diagnosis.

Characterization of marker chromosomes in patients with malignant myeloid disorders

We have used chromosome painting to characterize small marker chromosomes in a series of patients with malignant myeloid disorders [60]. Partial or complete loss of chromosome 7 occurs in all subgroups of AML and myelodysplastic syndromes (MDS) (see Appendix IX, Tables IX.1 and IX.5), in both adults and children. In all cases this chromosome abnormality predicts a poor response to treatment and short survival times. In a proportion of patients with apparent monosomy 7 there are additional uncharacterized marker chromosomes, especially small fragments or rings. We suspected that these small chromosomes may represent progressive deletion of one chromosome 7 homologue and evolution to true monosomy 7. We used FISH with whole-chromosome painting probes to investigate the origin of ring (r) or marker (mar) chromosomes in seven patients whose karyotype included -7. Hybridization to bone marrow metaphase cells was carried out using a chromosome 7 paint derived from Alu-PCR-amplified flow-sorted chromosomes 7, and purchased paints for chromosomes 5 and 18 (Cambio). The results are summarized in Table 10.5.

In patients 1–4 (Table 10.5) the ring chromosomes were confirmed to be of chromosome 7 origin. In patient 2, the ring chromosome was not highlighted by the chromosome 7 paint, but was later confirmed to contain only chromosome 7 centromeric material, by FISH with a chromosome 7-specific centromere probe. This abnormality has therefore been redefined as a centromeric fragment. Patients 5 and 6 (Table 10.5) were shown to have cells with true monosomy 7, as the ring and marker chromosomes in these cases failed to hybridize with the chromosome 7 paint. Hybridization of cells from patient 7 with a chromosome 5 and 7 paint identified the marker chromosome as having both chromosome 5 and chromosome 7 material present. In patients 3 and 4 FISH revealed selective loss of the r(7) chromosome in a proportion of cells. The value of chromosome painting in this study has been to confirm that in many cases the small marker chromosomes accompanying -7 in complex karyotypes are in fact derived from chromosome 7. The ability to characterize such marker chromosomes accurately may eventually lead to a redefinition of prognostic groups.

Case Study 10.1

Table 10.5 Complete and partial monosomy 7 in leukaemia studied by FISH.

Patient	Diagnosis	Chromosome abnormality	Revised abnormality
1	T-ALL/AML	add(7), -7, +r	dup(7), r(7)
2	RA	-7, +?r	der(7)cen
3	AML-M6 (previously RAEB)	-5, -6, -7, +r	-5, -6, r(7)
4	Myelofibrosis	-7, +r	r(7)
5	AML-M6	-7, -18, +r	-7, r(18)
6	AML-M2	-5, -7, -18, +mar1, +mar2	-5, -7, -18, + mar1, +mar2
7	AML-M2 previously RAEB)	-5, add(7), add(7), add(9), -12, -18, +mar	der(5)t(7;5;7), der(7), der(7), -12, -18

AML, acute myeloid leukaemia; T-ALL, T-cell acute leukaemia; RA, refractory anaemia; RAEB, refractory anaemia with excess blasts; r, ring; mar, marker chromosome; dup, duplication; der, derivative; add, additional unidentified material present. See Chapters 7 and 8 and Appendix IX for further information on chromosome aberrations associated with congenital abnormalities and cancer.

.....

Protocol 50 Nick translation

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.

- Using a commercial kit. A number of commercial kits are available for nick translation. We have found that the BRL Bionick kit (Life Technologies (Gibco-BRL)) gives good results because it has a modified DNase I concentration to give optimally sized fragments for *in situ* hybridization. A method for nick translation using this kit is given in Chapter 9, Protocol 44.
- Alternative protocol. The size of labelled DNA fragments is critical to the success of hybridization, with an average size of 300 bp (range of 100–500 bp) being most suitable. Larger fragments produce high background signals, as access to the target sequence is impeded. This method allows the fragment size to be controlled by titration of DNase I concentration, and is also considerably cheaper than commercial kits.

Materials

- Purified DNA: whole chromosome library, PCR-amplified chromosomes
- 10×nick translation buffer: 0.5 M Tris-HCl (pH 7.5), 50 mM MgCl₂, 0.5 mg ml⁻¹ nuclease-free BSA
- biotin-16-dUTP (1 mM), digoxigenin-11-dUTP (1 mM) (Boehringer Mannheim)
- DTT (100 mM) (Sigma)

- dNTP mix: 0.5 mM each dATP, dCTP, dGTP, and 0.1 mM dTTP (Boehringer)
- DNase I ($20 \mu\text{g } \mu\text{l}^{-1}$) (Amersham International)
- DNase I dilution buffer: 50% glycerol, 0.15 M NaCl, 20 mM sodium acetate (pH 5.0)
- DNA polymerase I ($3.5 \text{ U } \mu\text{l}^{-1}$) (Amersham)
- Sephadex G50 (equilibrated in 10 mM Tris, 1 mM EDTA, pH 8.0) (Pharmacia)
- 1-ml syringes
- select B columns (CP Laboratories)
- *Escherichia coli* tRNA (Boehringer)
- salmon sperm DNA (Sigma) sonicated to an average size of 500 bp
- TE: 10 mM Tris-HCl (pH 7.5), 1 mM EDTA
- ϕX174 *Hae*III size marker (BRL Life Technologies)
- 3 M sodium acetate, pH 5.2
- agarose (UltraPure grade, BRL Life Technologies)
- ethidium bromide (Sigma)
- absolute ethanol

Method

- 1 To a sterile 1.5-ml microcentrifuge tube add (in order):
 - 1 μg probe DNA;
 - 1.2 μl biotin-16-dUTP (1 mM) or digoxigenin-11-dUTP (1 mM);
 - 5 μl 10 \times nick translation buffer;
 - 5 μl DTT (100 mM);
 - 5 μl dNTP mix;
 - sterile distilled water to final volume of 50 μl ;
 - 3 μl DNA polymerase I;
 - 3–6 μl DNase I.
- 2 Incubate for 90 min at 15 °C.
- 3 Stop the reaction by placing the tube on ice.
- 4 Check the probe fragment size range by running a 5 μl aliquot on a 2% agarose gel with ϕX174 *Hae*III size markers. The desired size range for optimal hybridization is 100–500 bp. If the probe fragment size is too large, extra DNase I can be added and the reaction mixture reincubated at 15 °C for a further 30–60 min.
- 5 Purify the labelled probe through a Sephadex G50 spin column in a 1-ml syringe to remove unincorporated nucleotides as follows:
 - Load Sephadex G50 (equilibrated in 10 mM Tris, 1 mM EDTA, pH 8.0) into a 1-ml syringe (sealed at one end with filter wool) and pack to a height of 10 cm by centrifugation at 1200 *g* for 3 min.
 - Add the labelled probe (50 μl per column) to the top of the column and collect the purified eluate into a 1.5-ml microcentrifuge tube by centrifugation at 1200 *g* for 3 min.
 - To the purified eluate add 50 μg *E. coli* tRNA, 50 μg salmon sperm

DNA, 0.1 vol. of 3 M sodium acetate (pH 5.2) and 2 vols of ice-cold ethanol.

- Precipitate the DNA at -30°C overnight then centrifuge to recover the pellet. Remove the supernatant and dry the pellet before resuspending in 50 μl of TE (pH 7.5). Purified, labelled probes are stable when stored at -30°C for several years.

Alternatively, the probe can be purified using a commercially available column such as Select B, specifically designed for biotin-labelled probes.

To titrate DNase I concentration:

- Make up a 1 mg ml^{-1} stock solution, then a 2.5 $\text{ng } \mu\text{l}^{-1}$ working solution. Each new batch of working solution is tested as follows: set up a standard nick translation reaction (without DNA polymerase I or dNTPs) using 1 μg of DNA which is known to cut in the desired range and increasing amounts of 2.5 $\text{ng } \mu\text{l}^{-1}$ DNase (e.g. 3, 5 and 6 μl). Stop the reaction by placing the tubes on ice. Check the size range on a 2% agarose gel with $\phi\text{X174 HaeIII}$ as a size marker.

.....

Protocol 51 Alu-PCR amplification of flow-sorted chromosomes

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- 250–500 flow-sorted chromosomes
- Alu-BK33 primer: 5'-CTGGGATTACAGGCGTGAGC-3' [38]
- dNTPs (Pharmacia, UltraPure grade)
- PCR buffer: 1.5. mM MgCl_2 , 50 mM KCl, 10 mM HCl, 0.1% (w/v) Triton X-100, 0.01% (w/v) gelatin (pH 8.9)
- *Taq1* polymerase (Boehringer)
- mineral oil (Sigma)
- sterile distilled water
- programmable thermal cycler
- agarose, ethidium bromide, gel electrophoresis apparatus, UV transilluminator
- molecular weight marker ($\phi\text{X174 HaeIII}$)
- set of micropipettes (kept for PCR only) and sterile tips
- 0.5-ml PCR tubes

Method

- 1 Between 250 and 500 chromosomes can be sorted directly into 0.5-ml microcentrifuge tubes containing 100 μl aliquots of 10 $\mu\text{g ml}^{-1}$ Alu-BK33 primer, 200 μM each of dNTPs, PCR buffer, 25 U ml^{-1} *Taq1* polymerase.

- 2 Prepare a negative control containing the appropriate amount of sheath fluid but no chromosomes and a positive control containing 2.5 pg genomic DNA in the same way.
- 3 Mix the reagents gently and overlay with 100 µl mineral oil before carrying out the following PCR programme in a DNA thermal cycler:
- 1 min at 95 °C (initial denaturation)
 - 35 cycles of:
 - 30 s at 95 °C
 - 1 min at 55 °C
 - 4 min at 68 °C
- with the final extension time lengthened to 10 min.
- 4 Remove the oil and run a 10-µl aliquot of amplified products (from the control samples as well as tests) on a 1.2% agarose gel with *HaeIII*-cleaved φX174 DNA marker to check the amount and size of the amplified products. The products will be seen as a smear running the length of the φX174 ladder. At this stage the amplified products can be stored at –20 °C until required.
- 5 Purify the amplified products by gel filtration through a Sephadex G50 spin column (Protocol 50). Measure the DNA concentration accurately on a fluorimeter (this is usually about 20 ng µl⁻¹).
- 6 Carry out an ethanol precipitation with 0.1 vol. 3 M sodium acetate (pH 5.2) and 2 vols ethanol at –30 °C for at least 2 h (overnight is preferable).
- 7 Centrifuge in a microcentrifuge for 15 min at 4 °C to recover the DNA pellet. Remove the supernatant and dry the pellet (in a vacuum desiccator or air-dry). Resuspend the DNA pellet in sterile distilled water at a suitable concentration for labelling by nick translation (see Protocol 50 and Chapter 9).

.....

Protocol 52 DOP-PCR amplification of flow-sorted normal and abnormal chromosomes

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- flow-sorted chromosomes (approximate concentration, 500 per µl)
- 2µl PCR buffer: 10 mM MgCl₂, 100 mM KCl, 20 mM Tris-HCl (pH 8.4), 0.2 mg ml⁻¹ gelatin
- dNTP mix 2 mM each dATP, dCTP, dGTP, dTTP
- 6-MW primer: 5'-CCGACTCGAGNNNNNNATGTGG-3' [9]
- *Taq*1 polymerase (Boehringer Mannheim)
- biotin-16-dUTP (1 mM) or digoxigenin-11-dUTP (1 mM) (Boehringer Mannheim)

Method

FIRST ROUND OF AMPLIFICATION

- 1** Combine in a sterile 0.5-ml microcentrifuge tube: 500 flow-sorted chromosomes, 50 μ l 2 \times PCR buffer, 10 μ l dNTP mix, 6.6 μ l 6-MW primer (30 μ M), 1.25 U Taq1 polymerase, water to a final volume of 100 μ l.

All solutions, microcentrifuge tubes, and tips should be autoclaved and kept sterile. The reagents should be added in a sterile laminar flow hood to avoid the possibility of contamination. As a further precaution against contamination all the above reagents except chromosomes and Taq1 polymerase can be sterilized by exposure to short-wave UV irradiation (10 min on a UV transilluminator) prior to the first round of amplification.

- 2** Prepare positive (2.5 pg genomic DNA instead of flow-sorted chromosomes) and negative (all of the above reagents except DNA) controls in the same way.
- 3** Overlay the reaction mixture with 100 μ l mineral oil and carry out the following programme in a PCR thermal cycler:
 - 10 min at 93 °C (initial denaturation step);
 - 5 cycles of:
 - 1 min at 94 °C
 - 1.5 min at 30 °C
 - 3 min at 30–72 °C (transition)
 - 3 min at 72 °C
 - 35 cycles of:
 - 1 min at 94 °C
 - 1 min at 62 °C
 - 3 min at 72 °C
 - with an additional 1 s per cycle and a final extension time of 10 min.
- 4** Remove the mineral oil. At this stage the amplified products can be stored at –20 °C until required. Remove a 10- μ l aliquot of amplified products (from the control tubes as well) and run on a 1.2% agarose gel with ϕ X174 markers to check the success of the amplification. There should be no amplification of the negative control (see Fig. 10.1).

SECOND ROUND OF AMPLIFICATION AND BIOTIN LABELLING

- 5** To a new sterile microcentrifuge tube add the following: 5 μ l of the amplified products from round 1, 25 μ l 2 \times PCR buffer, 5 μ l dNTP mix (as before), 3.3 μ l 6-MW primer (30 μ M), 0.625 U Taq 1 polymerase, 12 μ l biotin-16-dUTP or digoxigenin-11-dUTP (1 mM).
- 6** Mix well, overlay with 50 μ l mineral oil and place in a DNA thermal cycler with the following program:
 - 10 min at 93 °C (initial denaturation step);
 - 25 cycles of:

- 1 min at 94 °C
 - 1 min at 62 °C
 - 3 min at 72 °C
 - with a final extension time of 10 min.
- 7** Remove the mineral oil. Run 10 µl labelled products on a 1.2% agarose gel to check the size range. If too large re-cut with 5 µl DNase I for 30–60 min.
- 8** Purify the labelled DNA through a spin column (see Protocol 50). Measure the DNA concentration of the purified labelled DNA in a fluorimeter. The concentration is usually 20–50 ng µl⁻¹. Ethanol precipitate the labelled DNA as described in Protocol 50. The probe is now ready for use as a chromosome paint in CISSH experiments.

.....

Protocol 53

Competitive (or chromosomal) in situ suppression hybridization

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- slide containing metaphase chromosomes (see Chapter 7 for protocols)
- human Cot-1 DNA (BRL Life Technologies)
- sodium acetate (3 M)
- hybridization buffer: 50% (w/v) formamide, 10% (w/v) dextran sulphate, 1% (v/v) Triton X- 100, 2×SSC (pH 7.0)
- ethanol series (70%, 95% and absolute)
- formamide (Fluka)
- 50% dextran sulphate
- 20×SSC : 1×SSC = 150 mm sodium chloride, 15 mm sodium citrate (pH 7.0)
- RNase A (10 mg ml⁻¹; Sigma) boiled for 10 min to remove contaminating DNase
- formaldehyde (40% w/v)
- coverslips, Coplin jars, rubber cement, water baths

Method

PREPARATION OF PROBE AND COMPETITION WITH UNLABELLED GENOMIC DNA

- 1** Prepare labelled probes by nick translation with biotin or digoxigenin as in Protocol 50, or by DOP-PCR as in Protocol 52.
- 2** Prior to hybridization add the following to a 1.5-ml microcentrifuge tube:

- labelled probe (100 ng if from a whole chromosome library or DOP-PCR amplified flow-sorted chromosomes; 400 ng Alu-PCR if from flow-sorted chromosomes);
 - unlabelled competitor DNA (10 µg for whole chromosome libraries or DOP-PCR flow-sorted chromosomes; 40 µg for Alu-PCR flow-sorted chromosomes).
- 3** Precipitate with 0.1 vol. 3 M sodium acetate and 2 vols of ice-cold ethanol for 1 h at -70°C . Centrifuge in a microcentrifuge to pellet the DNA for 15 min, discard the supernatant, dry the pellet and resuspend in 11–15 µl hybridization buffer (50% (v/v) formamide, 10% (w/v) dextran sulphate, 1% (v/v) Triton X100, and 2×SSC (pH 7.0)).
 - 4** Heat the mixture of probe and competitor to $70\text{--}95^{\circ}\text{C}$ for 5 min, then chill on ice, before incubating at 37°C for 15 min to allow partial reannealing.

PRETREATMENT OF CHROMOSOMES

- 5** Treat the chromosomal DNA on slides with RNase (100 µg ml⁻¹ in 2×SSC, boiled to remove contaminating DNase at 37°C for 1 h). Wash slides in three changes of 2×SSC, and dehydrate through an ethanol series (50%, 75%, 95%, and absolute, 3 min each).
- 6** The following postfixation steps help access of probe to the target DNA, and are particularly important for interphase analysis. Immerse slides in the following solutions:
 - PBS containing 50 mM MgCl₂ for 5 min;
 - PBS/50 mM MgCl₂/1% formaldehyde for 10 min;
 - PBS for 5 min.
- 7** Dehydrate through alcohol series (10%, 50%, 70%, 95% and absolute ethanol).
- 8** Denature chromosomal DNA (metaphase or interphase cells) by immersing slides in 70% (v/v) formamide, 2×SSC (pH 7.0) at 75°C for 3–5 min, followed by dehydration through a cold ethanol series (70%, 95% and absolute, 3 min each).

HYBRIDIZATION OF PROBE TO CHROMOSOMES

- 9** Air-dry and place the previously annealed probe mixture on the slide and cover with a 22×22 mm glass coverslip. Seal the edges with rubber solution and place the slides in a sealed box in a water bath at 37°C overnight or for up to 96 h.
 - 10** After hybridization, remove the rubber solution and immerse the slides in 2×SSC for 5 min to float off the coverslips.
 - 11** Posthybridization washes (all at 42°C):
 - three 5 min washes in 50% (v/v) formamide, 2×SSC (pH 7.0);
 - three 5 min washes in 2×SSC (pH 7.0).
- Caution: Formamide is toxic by inhalation, in contact with skin and if*

swallowed. It may cause birth defects. All steps involving formamide, particularly hot formamide, should be carried out in a fume hood.

Alternatively, to avoid the use of hot formamide solutions the following posthybridization washes give equally good results:

- three 5-min washes in $2\times$ SSC at room temperature (with agitation);
- two 20-min washes in $0.1\times$ SSC at 65°C followed by;
- one 5-min wash in $0.1\times$ SSC at room temperature (with agitation).

.....

Protocol 54 Detection of hybridized labelled probe

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- fluorescence microscope (epifluorescence illumination) with suitable fluorescence objectives and filter sets
- blocking solution (SSCT-BSA): 3% (w/v) BSA in $4\times$ SSC, 0.05% (v/v) Triton X-100
- wash solution (SSCT): $4\times$ SSC, 0.05% (v/v) Triton X-100
- avidin-DCS-FITC (1 mg ml^{-1}) (Vector Laboratories)
- biotinylated antiavidin (Vector Laboratories)
- DAPI (4,6-diamidino-2-phenylindole) (Sigma)
- propidium iodide (Sigma)
- Citifluor AF1 mountant (Citifluor)
- Vectashield mountant (Vector Laboratories): a self-prepared mixture containing 22 mg 1,4-diazobicyclo (2.2.2.) octane (DABCO) in 1 ml 20 mM NaHCO_3 (pH 8.0), 75% glycerol (61), or 10 mg ml^{-1} *p*-phenylenediamine in PBS mixed 1:9 with glycerol and adjusted to pH 8.0 with 0.5 M carbonate–bicarbonate buffer (pH 9.0)
- avidin–DCS–Texas red (Vector Laboratories)
- monoclonal antidigoxigenin antibody (Sigma)
- FITC-labelled rabbit antimouse immunoglobulin (Sigma)
- monoclonal FITC-labelled antirabbit immunoglobulin (Sigma)

Method

- 1** Incubate slides in blocking solution for 15–30 min at room temperature to block nonspecific protein-binding sites.
- 2** Wash slides in wash solution (SSCT) before adding one of the following detection reagents.
- 3a** *Biotin-labelled probes* Biotinylated probes are detected with avidin–DCS (cell sorter grade) conjugated to fluorescein

isothiocyanate (FITC). Detection reagents are diluted in blocking solution (filtered through a 22- μm syringe filter) and washes carried out in SSCT. All incubations with detection reagents are carried out in a humidified chamber at 37 °C.

- Place 100 μl avidin–DCS–FITC (5 $\mu\text{g ml}^{-1}$) on a slide, cover with a 50 mm \times 24 mm coverslip and place in a humid chamber at 37 °C for 30 min.
- Wash in three changes of SSCT.
- Add 100 μl biotinylated antiavidin (5 $\mu\text{g ml}^{-1}$) per slide and incubate for 20 min at 37 °C.
- Wash three times in SSCT.
- Add 100 μl avidin–FITC and incubate for 20 min at 37 °C.

Finally, slides are washed once in SSCT, followed by two 5 min washes in PBS (pH 7.0), and dehydrated through a ethanol series (70%, 95% and absolute, 3 min each). After air-drying, slides are mounted in Citifluor AF1 or Vectashield mountant containing 0.5 $\mu\text{g ml}^{-1}$ propidium iodide (PI) as counterstain. Seal coverslips with nail varnish and store slides at 4 °C in the dark. A mixture of DAPI (1.5 $\mu\text{g ml}^{-1}$) and PI (0.75 $\mu\text{g ml}^{-1}$) in the mountant can be used to identify chromosomes by producing a G-banding pattern when viewed under UV filters and an R-banding pattern when viewed under the green filter set.

3b Digoxigenin-labelled probes Probes labelled with digoxigenin are detected by incubation with:

- 1st layer: monoclonal antidigoxigenin antibody (1.5 μl in 1 ml blocking solution) for 30 min at 37 °C, followed by;
- 2nd layer: rabbit antimouse-FITC (1 μl in 1 ml blocking solution) for 30 min at 37 °C;
- 3rd layer: monoclonal antirabbit-FITC (10 μl in 1 ml blocking solution).

Wash three times in wash solution between each layer as for single-colour detection. The final washing and dehydration steps are carried out as for biotin detection.

3c Dual-colour detection Make up the following antibody dilutions in 1 ml blocking solution:

- 1st layer: 8 μl avidin–Texas red (2.5 mg ml^{-1} stock) plus 1.5 μl monoclonal antidigoxigenin antibody.
- 2nd layer: 10 μl biotin antiavidin (0.5 mg ml^{-1}) stock plus 1 μl rabbit antimouse-FITC.
- 3rd layer: 8 μl avidin–Texas red plus 10 μl monoclonal antirabbit-FITC.

Incubate slides in each antibody layer (100 μl under a 24 \times 50 mm coverslip) for 30 min at 37 °C in a moist chamber.

All reagents are diluted in SSCT-BSA (blocking solution) with three 3-min washes between each incubation as for single-colour detection. The final washing and dehydration steps are carried out as for biotin detection. Mount in antifade medium containing only DAPI counterstain.

.....

Troubleshooting

(See also Chapter 9)

Amplification of negative control after first round DOP-PCR

Because of the low stringency of 1st round amplification with the MW6 primer, all reagents, tubes and tips must be autoclaved, kept sterile, and kept only for PCR. We always add the reagents in a sterile laminar flow hood to avoid the possibility of contamination.

Cell lost from slide

Handle slides carefully at all stages, especially during removal of coverslips. Agitation during posthybridization washes should be carried out on a rocking platform set at minimum speed.

Poor chromosome morphology

If chromosomes look puffy they may have been over-denatured: always check temperature of denaturing solution inside the Coplin jar.

No hybridization signal

This may be due to:

- *Insufficient probe DNA in the hybridization mix. DNA concentration of any new probe should be measured accurately on a fluorimeter or on a gel against a range of concentrations of uncut lambda DNA.*
- *Inadequate denaturation of probe and/or chromosomes.*
- *Probe fragment size too small. Always check the labelled fragment size on a 2% (1.2% for PCR products) gel with ϕ X174 HaeIII size marker. The optimum fragment size is 100–500 bp.*

High background

High background with strong specific signal may be due to:

- *Low stringency of hybridization or posthybridization washes. The stringency of hybridization can be increased by either increasing the hybridization temperature, increasing the formamide concentration of the hybridization mix and/or posthybridization washes to 60%, or decreasing the SSC concentration to 0.1% in the posthybridization washes.*
- *Incomplete competition. Increase the Cot-1 DNA concentration. Cot-1 DNA is already present in large excess so any increase should be substantial (up to 10-fold). A brightly fluorescent background signal all over the slide, obscuring any specific signal, occurs when the labelled probe fragments are too large. If labelled probe is > 500 bp it should be re-cut with DNase I. High background of this type may also be caused by insufficient blocking with BSA.*

.....

References

1 Pinkel, D., Landegent, J., Collins, C. *et al.* (1988) Fluorescence *in situ* hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc Natl. Acad. Sci. USA* **85** (23), 9138–9142.

2 Lichter, P., Cremer, T., Borden, J. *et al.* (1988) Delineation of individual human chromosomes in metaphase and interphase cells by *in situ* suppression hybridization using recombinant DNA libraries. *Hum. Genet.* **80** (3), 224–234.

3 Cremer, T., Lichter, P., Borden, J. *et al.* (1988) Detection of chromosome aberrations in metaphase and interphase tumor cells by *in situ* hybridization using chromosome-specific library probes. *Hum. Genet.* **80** (3), 235–246.

4 Kievits, T., Devilee, P., Wiegant, J. *et al.* (1990) Direct nonradioactive *in situ* hybridization of somatic cell hybrid DNA to human lymphocyte chromosomes. *Cytometry* **11** (1), 105–109.

5 Pinkel, D., Straume, T. & Gray, J.W. (1986) Cytogenetic analysis using quantitative, high sensitivity fluorescence hybridization. *Proc. Natl. Acad. Sci. USA* **85** (9), 2934–2938.

6 Nizetic, D., Zehetner, G., Monaco, A.P. *et al.* (1991) Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: their potential use as reference libraries. *Proc Natl. Acad. Sci. USA* **88** (8), 3233–3237.

7 Collins, C., Kuo, W.L., Segraves, R. *et al.* (1991) Construction and characterization of plasmid libraries enriched in sequences from single human chromosomes. *Genomics* **11** (4), 997–1006.

8 Fuscoe, J.C., Collins, C.C., Pinkel, D. & Gray, J.W. (1989) An efficient method for selecting unique-sequence clones from DNA libraries and its application to fluorescent staining of human chromosome 21 using *in situ* hybridization. *Genomics* **5**, 100–109.

9 Telenius, H., Carter, N.P., Bebb, C.E. *et al.* (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* **13** (3), 718–725.

10 Telenius, H., Pelmear, A.H., Tunnacliffe, A. *et al.* (1992) Cytogenetic analysis by chromosome painting using DOP-PCR amplified flow-sorted chromosomes. *Genes Chrom. Cancer* **4** (3), 257–263.

11 Carter, N.P., Ferguson-Smith, M.A., Perryman, M.T. *et al.* (1992) Reverse chromosome painting: a method for the rapid analysis of aberrant chromosomes in clinical cytogenetics. *J. Med. Genet.* **29** (5), 299–307.

12 Rack, K.A., Harris, P.C., MacCarthy, A.B. *et al.* (1993) Characterization of three de novo derivative chromosomes 16 by ‘reverse chromosome painting’ and molecular analysis. *Am. J. Hum. Genet.* **52**, 987–997.

13 Meltzer, P.M., Guan, X.-Y., Burgess, A. & Trent, J.M. (1992) Rapid generation of region-specific probes by microdissection and their application. *Nature Genet.* **1**, 24–28.

14 Ohta, T., Soejima, H., Fukushima, Y. *et al.* (1993) The origin of cytologically unidentifiable chromosome abnormalities: six cases ascertained by targeted chromosome-band painting. *Hum. Genet.* **92**, 1–5.

15 Zhang, J., Trent, J.M. & Meltzer, P.S. (1993) Rapid isolation and characterization of amplified DNA by chromosome microdissection: identification of IGF1R amplification in malignant melanoma. *Oncogene* **8**, 2827–2831.

16 Guan, X.-Y., Meltzer, P.S., Dalton, W.S. & Trent, J.M. (1994) Identification of cryptic sites of DNA sequence amplification in human breast cancer by chromosome microdissection. *Nature Genet.* **8**, 155–161.

17 Nederlof, P.M., Robinson, D., Abuknesha, R. *et al.* (1989) Three-color fluorescence *in situ* hybridization for the simultaneous detection of multiple nucleic acid sequences. *Cytometry* **10** (1), 20–27.

18 Nederlof, P.M., van der Flier, S., Wiegant, J. *et al.* (1990) Multiple fluorescence *in situ* hybridization. *Cytometry* **11** (1), 126–131.

19 Ried, T., Baldini, A., Rand, T.C. & Ward, D.C. (1992) Simultaneous visualization of seven different DNA probes by *in situ* hybridization using combinatorial fluorescence and digital imaging microscopy. *Proc Natl. Acad. Sci. USA* **89** (4), 1388–1392.

20 Ried, T., Landes, G. & Dackowski, W. (1992) Multicolor fluorescence *in situ* hybridization for the simultaneous detection of probe sets for chromosomes 13 (18), 21, x and Y in uncultured amniotic cells. *Hum. Mol. Genet.* **1** (5), 307–313.

21 Dauwerse, J.G., Wiegant, J., Raap, A.K., Breuning, M.H. & van Ommen, G.J.B. (1992) Multiple colors by fluorescence *in situ* hybridization using ratio-labelled DNA probes create a molecular karyotype. *Hum. Mol. Genet.* **1** (8), 593–598.

22 Speicher, M.R., Ballard, S.G. & Ward, D.C. (1996) Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genet.* **12**, 368–375.

23 Schrock, E., du Manoir, S., Veldman, T. *et al.* (1996) Multicolor spectral karyotyping of human chromosomes. *Science* **273**, 494–497.

24 Popp, S., Jauch, A., Schindler, D. *et al.* (1993) A strategy for the characterization of minute chromosome rearrangements using multiple color fluorescence *in situ* hybridization with chromosome-specific DNA libraries and YAC clones. *Hum. Genet.* **92**, 527–532.

25 Lengauer, C., Speicher, M.R., Popp, S. *et al.* (1993) Chromosomal bar codes produced by multicolor fluorescence *in situ* hybridization with multiple YAC clones and whole chromosome painting probes. *Hum. Mol. Genet.* **2** (5), 505–512.

26 Vooijs, M., Yu, L.-C., Tkachuk, D. *et al.* (1993) Libraries for each human chromosome, constructed from sorter-enriched chromosomes by using linker-adaptor PCR. *Am. J. Hum. Genet.* **52**, 586–597.

27 Suijkerbuijk, R.F., van de Veen, A.Y., van Echten, J. *et al.* (1991) Demonstration of the genuine iso-12p character of the standard marker chromosome of testicular germ cell tumors and identification of further chromosome 12 aberrations by competitive *in situ* hybridization. *Am. J. Hum. Genet.* **48** (2), 269–273.

28 Nelson, D.L., Ledbetter, S.A., Corbo, L. *et al.* (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. USA* **86** (17), 6686–6690.

29 Britten, R.J., Baron, W.F., Stout, D.B. & Davidson, E.H. (1988) Sources and evolution of human Alu repeated sequences. *Proc Natl. Acad. Sci. USA* **85** (13), 4770–4774.

- 30 Moyzis, R.K., Torney, D.C., Meyne, J. *et al.* (1989) The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics* **4**, 273–289.
- 31 Ledbetter, S.A., Nelson, D.L., Warren, S.T. & Ledbetter, D.H. (1990) Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. *Genomics* **6** (3), 475–481.
- 32 Lichter, P., Ledbetter, S.A., Ledbetter, D.H. & Ward, D.C. (1990) Fluorescence *in situ* hybridization with Alu and L1 polymerase chain reaction probes for rapid characterization of human chromosomes in hybrid cell lines. *Proc Natl. Acad. Sci. USA* **87** (17), 6634–6638.
- 33 Lengauer, C., Riethman, H. & Cremer, T. (1990) Painting of human chromosomes with probes generated from hybrid cell lines by PCR with Alu and L1 primers. *Hum. Genet.* **86** (1), 1–6.
- 34 Cotter, F.E., Hampton, G.M., Nasipuri, S. (1990) Rapid isolation of human chromosome-specific DNA probes from a somatic cell hybrid. *Genomics* **7** (2), 257–263.
- 35 Cotter, F.E., Nasipuri, S., Lam, G. & Young, B.D. (1989) Gene mapping by enzymatic amplification from flow-sorted chromosomes. *Genomics* **5** (3), 470–474.
- 36 Cotter, F.E., Das, S., Douek, E. *et al.* (1991) The generation of DNA probes to chromosome 11q23 by Alu PCR on small numbers of flow-sorted 22q-derivative chromosomes. *Genomics* **9** (3), 473–480.
- 37 Baldini, A. & Ward, D.C. (1991) *in situ* hybridization banding of human chromosomes with Alu-PCR products: a simultaneous karyotype for gene mapping studies. *Genomics* **9** (4), 770–774.
- 38 Korenberg, J.R. & Rykowski, M.C. (1988) Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell* **53** (3), 391–400.
- 39 Suijkerbuijk, R.F., Matthopoulos, D., Kearney, L. *et al.* (1992) Fluorescent *in situ* identification of human marker chromosomes using flow sorting and Alu element-mediated PCR. *Genomics* **13** (2), 355–362.
- 40 Joos, S., Scherthan, H., Speicher, M.R. *et al.* (1993) Detection of amplified DNA sequences by reverse chromosome painting using genomic tumour DNA as probe. *Hum. Genet.* **50** (4), 211–215.
- 41 Guan, X.-Y., Meltzer, P.S., Cao, J. & Trent, J.M. (1992) Rapid generation of region-specific genomic clones by chromosome microdissection: isolation of DNA from a region frequently deleted in malignant melanoma. *Genomics* **14**, 680–684.
- 42 Meltzer, P.S., Guan, X.-Y. & Trent, J. (1993) Telomere capture stabilizes chromosome breakage. *Nature Genet.* **4**, 252–255.
- 43 Zhang, J., Meltzer, P., Jenkins, R. *et al.* (1993) Application of chromosome microdissection probes for elucidation of BCR-ABL fusion and variant Philadelphia chromosome translocations in chronic myelogenous leukemia. *Blood* **81** (12), 3365–3371.
- 44 Guan, X.-Y., Zhang, H., Bittner, M. *et al.* (1996) Chromosome arm painting probes. *Nature Genet.* **12**, 10–11.
- 45 Tucker, J.D., Ramsey, M.J., Lee, D.A. & Minkler, J.L. (1993) Validation of chromosome painting as a biodosimeter in human peripheral lymphocytes following acute exposure to ionizing radiation *in vitro*. *Int. J. Radiat. Biol.* **64** (1), 27–37.
- 46 Manuelidis, L. (1990) A view of interphase chromosomes. *Science* **250**, 1533–1540.
- 47 Haaf, T. & Schmid, M. (1991) Chromosome topology in mammalian interphase nuclei. *Exp. Cell Res.* **192**, 325–332.
- 48 Kuo, W.-L., Tenjin, H., Segraves, R. *et al.* (1991) Detection of aneuploidy involving chromosomes 13, 18, or 21, by fluorescence *in situ* hybridization (FISH) to interphase and metaphase amniocytes. *Am. J. Hum. Genet.* **49**, 112–119.
- 49 Taylor, C., Patel, K., Jones, T. *et al.* (1992) Diagnosis of Ewing's sarcoma and peripheral neuroectodermal tumour based on the detection of t (11; 22) using fluorescence *in situ* hybridization. *Br. J. Cancer* **67**, 128–133.
- 50 Giovannini, M., Selleri, L., Biegel, J.A. *et al.* (1992) Interphase cytogenetics for the detection of the t (11; 22) (q24; q12) in small round cell tumours. *J. Clin. Invest.* **90**, 1911–1918.
- 51 Price, C.M., Kanfer, E.J., Colman, S.M. *et al.* (1992) Simultaneous genotypic and immunophenotypic analysis of interphase cells using dual-colour fluorescence: a demonstration of lineage involvement in polycythaemia vera. *Blood* **80** (4), 1033–1038.
- 52 Weber-Matthiesen, K., Winkemann, M., Muller-Hermelink, A. *et al.* (1992) Simultaneous fluorescence immunophenotyping and interphase cytogenetics: a contribution to the characterization of tumor cells. *J. Histochem. Cytochem.* **40** (2), 171–175.
- 53 Weber-Matthiesen, K., Deerberg, J., Muller-Hermelink, A. *et al.* (1993) Rapid immunophenotypic characterization of chromosomally aberrant cells by the new FICTION method. *Cytogenet. Cell Genet.* **63**, 123–125.
- 54 Nylund, S.J., Verbeek, W., Larramendy, M.L. *et al.* (1993) Cell lineage involvement in four patients with myelodysplastic syndrome and t (1; 7) or trisomy 8 studied by simultaneous immunophenotyping and fluorescence *in situ* hybridization. *Cancer Genet. Cytogenet.* **70**, 120–124.
- 55 Anastasi, J., Feng, J., Le Beau, M.M. *et al.* (1993) Cytogenetic clonality in myelodysplastic syndromes studied with fluorescence *in situ* hybridization: lineage, response to growth factor therapy, and clone expansion. *Blood* **81** (6), 1580–1585.
- 56 Liu, P., Siciliano, J., Seong, D. *et al.* (1993) Dual Alu polymerase chain reaction primers and conditions for isolation of human chromosome painting probes from hybrid cells. *Cancer Genet. Cytogenet.* **65**, 93–99.
- 57 Kallioniemi, A., Kallioniemi, O.P., Sudar, D. *et al.* (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumours. *Science* **258**, 818–821.
- 58 Wiegant, J., Wiesmeijer, C.C., Hoovers, J.M.N. *et al.* (1993) Multiple and sensitive fluorescence *in situ* hybridization with rhodamine-, fluorescein-, and coumarin-labelled DNAs. *Cytogenet. Cell Genet.* **63**, 73–76.
- 59 Jauch, A., Wienberg, J., Stanyon, R. *et al.* (1992) Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting. *Proc. Natl. Acad. Sci. USA* **89** (18), 8611–8615.
- 60 Gibbons, B., Lillington, D.M., Monard, S. *et al.* (1994) FISH studies to characterize complete and partial monosomy 7 in myeloid disorders. *Genes Chrom. Cancer* **10**, 244–249.

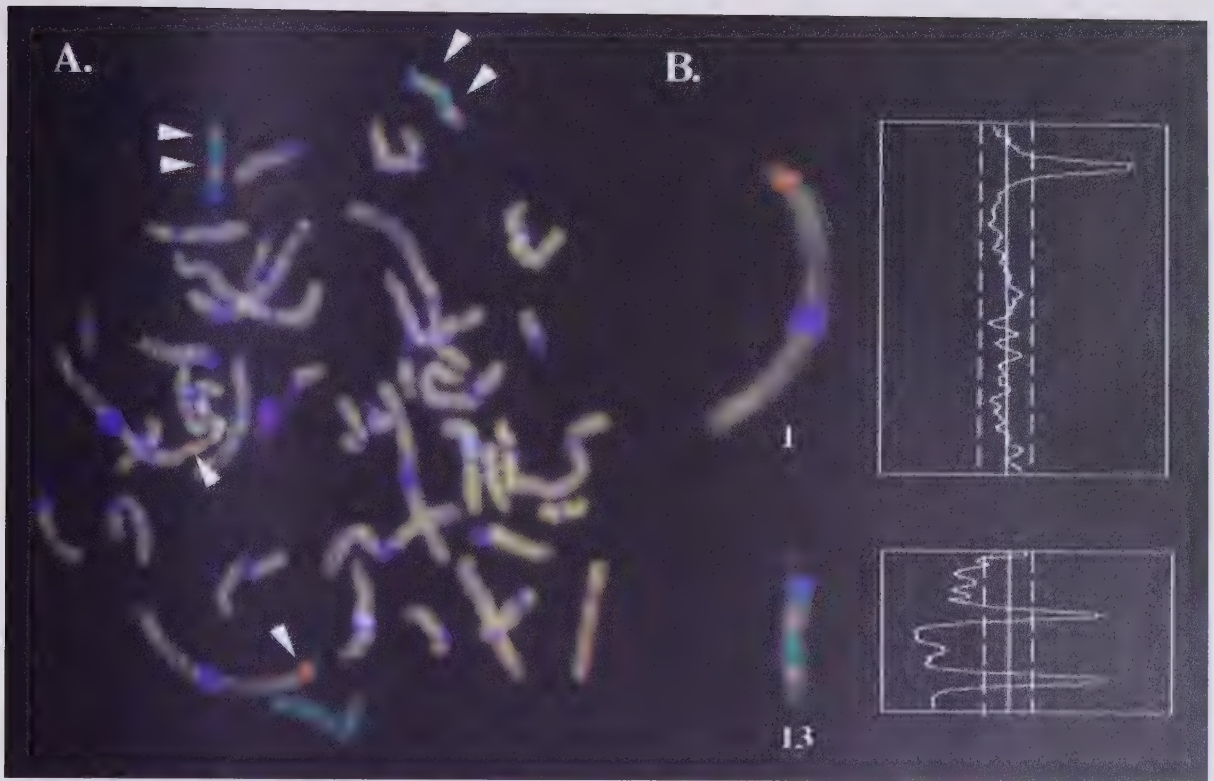


Plate 1 Comparative genomic hybridization analysis of the alveolar rhabdomyosarcoma case with double minutes shown in Fig. 8.5. A normal metaphase is shown in (A) following hybridization of tumour DNA (red) and normal control DNA (green). The arrows indicate the regions of amplification. The red/green fluorescence

ratios along the length of chromosomes 1 and 13 are shown in (B). A ratio outside the normal limits of 0.9–1.1 is indicative of a region-specific copy number change and in this case shows amplification at 1p36, 13q14 and 13q32 as well as loss of the whole of chromosome 13.



Plate 2 Interphase FISH analysis using tumour touch imprints of the alveolar rhabdomyosarcoma case with double minutes shown in Fig. 8.5. This shows a nucleus hybridized to a cosmid 5' of the *FKHR* gene (red), a *PAX7*

cosmid (green), and a 3' *FKHR* cosmid (blue). These markers are shown together in (A) and individually in (B), (C), and (D). This indicates co-amplification of all three probes.

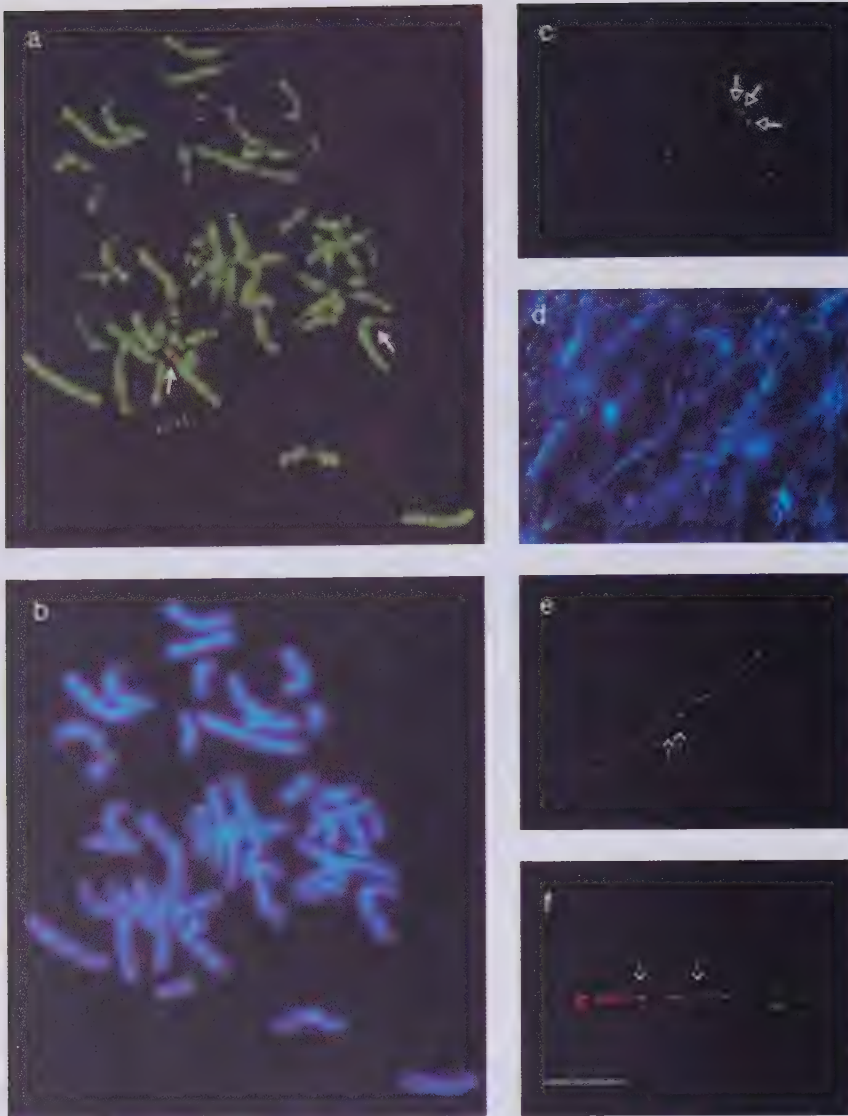


Plate 3 Examples for mapping and ordering of DNA probes by fluorescence *in situ* hybridization (FISH) (see Chapter 9). (a) Localization of a cosmid probe on replication G-banded metaphase chromosomes. The biotinylated probe pAKR4705, derived from the HLA class II region, was detected with avidin-Texas red with one round of signal amplification. The banding pattern was obtained by detecting BrdU, which was incorporated into the chromosomal DNA during the second half of the S-phase, with an FITC-conjugated anti BrdU antibody. Hybridization signals are visible on chromosome 6 in band 6p21.3 (arrows). (b) Identification of the chromosomes is facilitated using the DAPI staining, as the centromeres are not visible in the replication banding pattern. (c) Ordering of probes in interphase nuclei. Three cosmid probes from the HLA class II region were hybridized simultaneously to interphase nuclei. B51 was labelled with biotin and detected with Texas red, O27 was labelled with digoxigenin and detected with FITC and LH1 was labelled and detected with both systems,

resulting in a yellow colour when viewed using a dual band pass filter. The order red-green-yellow (arrows) suggests the probe order B51-O27-LH1. The distances are 300 kb between B51 and O27 and 600 kb between O27 and LH1 (16). (d) Chromatin released with sodium hydroxide and stained with DAPI is visible as a network of chromatin fibres. (e) and (f) show hybridization of two pairs of cosmids on sodium hydroxide released chromatin. In each case one probe was detected in green, the second in red. (e) Hybridization of two nonoverlapping cosmids clearly shows a gap between the two probe signals (arrows) while (f) shows the signals of two overlapping cosmids. The region of overlap is indicated by arrows, the bar represents 10 μ m. Photographs (a)–(c) were taken directly from the microscope. Pictures (d)–(f) were obtained from images captured with a cooled CCD camera (Photometrics), Pseudocolouring and merging of images was performed with computer software developed by T. Rand and D. Ward (Yale University, New Haven, CT).



Plate 4 FISH of whole chromosome painting probes produced by Alu-PCR amplification of flow-sorted chromosomes 1, 2, 3, and 4 (see Chapter 10). Normal chromosomes were purified by dual beam FACS analysis as described in Chapter 12. Each purified chromosome fraction was amplified by Alu-PCR as described in Chapter 10, Protocol 2, labelled with biotin and hybridized to normal metaphase chromosomes. Hybridized, labelled probe was detected with avidin

FITC, and the slides analysed by confocal laser scanning microscopy (Chapter 13). In each case the unlabelled chromosomes were counterstained with propidium iodide (red) and the labelled chromosomes 1 (A), 2 (B), 3 (C) and 4 (D) appear yellow. Note the R-banded pattern, particularly in (A) and (B) and the characteristic lack of centromere sequences detected by Alu-PCR derived paints.



Plate 5 FISH of whole chromosome painting probes produced by flow sorting of abnormal chromosomes (see Chapter 10). (A) Alu-PCR amplification of flow-sorted derivative chromosome 14 from the Daudi cell line, carrying the $t(8;14)(q24;q32)$. The der(14) paint highlights the chromosome 14 long arms (arrows), as well as the terminal region of chromosome 8 (arrowheads). (B) DOP-PCR amplification of the derivative chromosome



18 from the cell DOHH2 line carrying a $t(8;14;18)(q24;q32;q21)$. The der(18) paint highlights the majority of both chromosomes 18 (arrows) and a small segment on the terminal region of chromosome 14 (arrowheads). Photograph (B) courtesy of D. Lillington, Medical Oncology Unit, St. Bartholomew's Hospital, London.

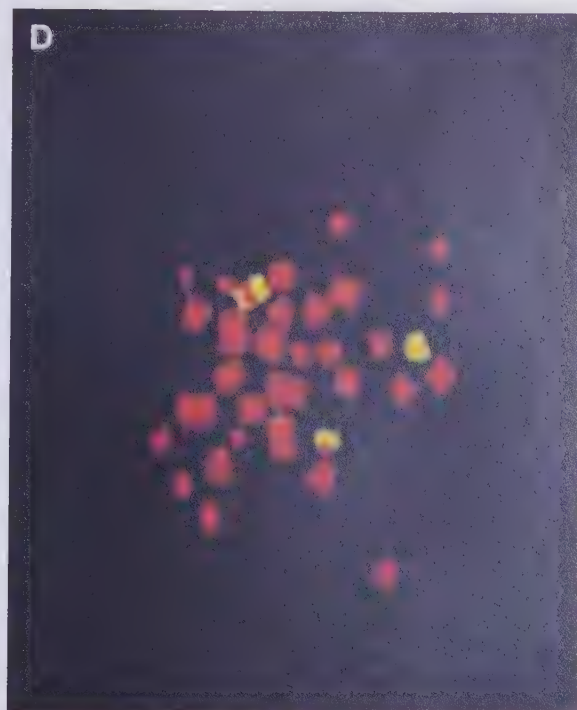
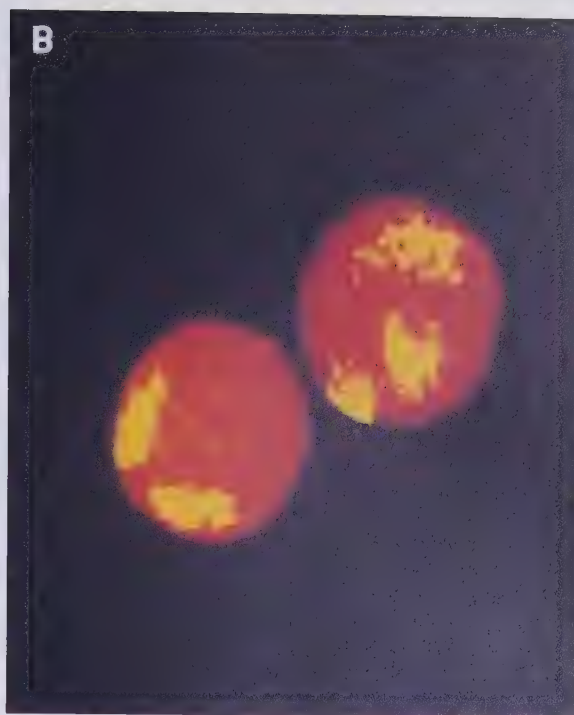
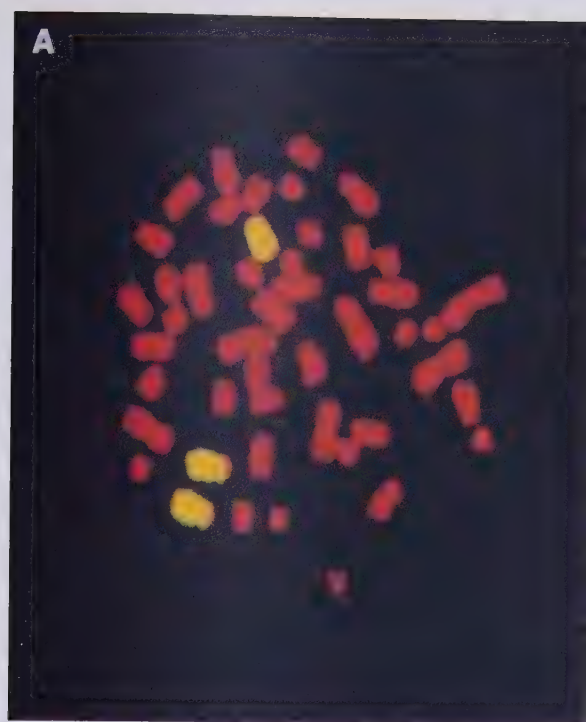


Plate 6 Detection of numerical and structural chromosome abnormalities with chromosome painting probes (see Chapter 10). (A) and (B) show FISH with a whole chromosome 8 paint (Cambio) to bone marrow cells from an AML patient: (A) a trisomy 8 metaphase cell and (B) two interphase cells showing two and three painting signals respectively. (C) and (D) show the characterization of marker chromosomes in leukaemic

patients with (C) a chromosome 5 paint (Cambio) and (D) an Alu-PCR-derived chromosome 7 paint. In (C) both the normal chromosome 5 (arrow) and the marker (arrowhead) are highlighted by the paint. In (D) three chromosomes have some chromosome 7 material present, but it is not possible to fully characterize any of these abnormal chromosomes using the chromosome 7 paint alone.



Plate 7 Micro-FISH analysis of DOP-PCR amplified material from the microdissection of chromosome region 6q26-27 (see Chapter 11). Micro-FISH was performed on normal metaphases. Arrows indicate FITC signal obtained from the biotin-labelled probe derived from the microdissected region (6q26-27). A biotin-labelled chromosome 6 centromere probe (Oncor) was also used to allow identification of chromosome 6.

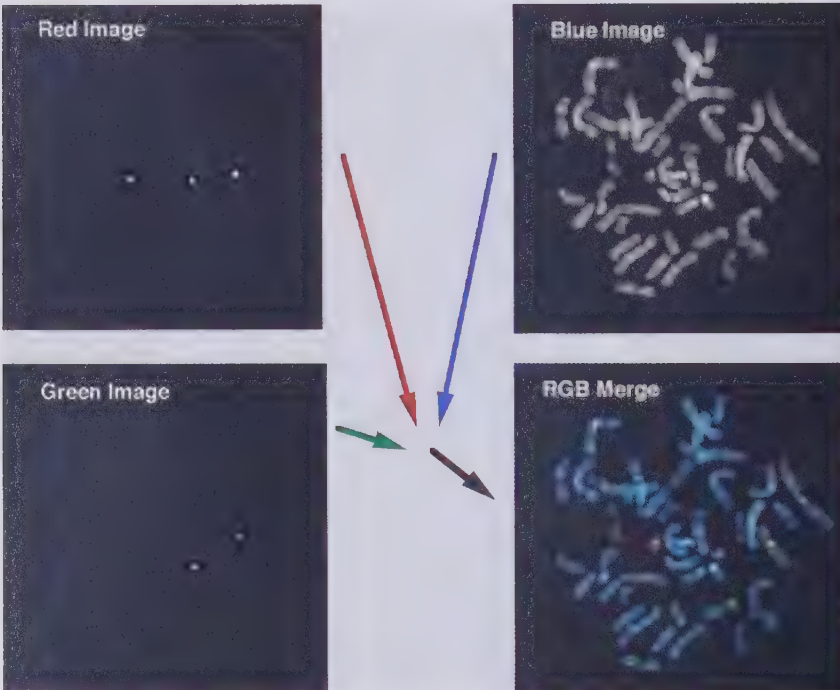


Plate 8 Colour merge of images acquired using a Photometrics KAF 1400 cooled CCD camera and SmartCapture software (Digital Scientific) (see Chapter 13). Blue image, DAPI counterstain; red image, signals from a chromosome 21 cosmid contig and an X-chromosome centromeric probe, fluorochrome Texas Red; green image, X-chromosome centromeric and Y chromosome centromeric probes, fluorochrome FITC. The dual-labelled X probe produces a yellow signal in the merged image. The blue plane of the RGB image (DAPI counterstain) was lightened using Adobe Photoshop for reproduction on a dye-sublimation colour printer. Original data provided by Yunling Zheng, Department of Pathology, University of Cambridge, UK.

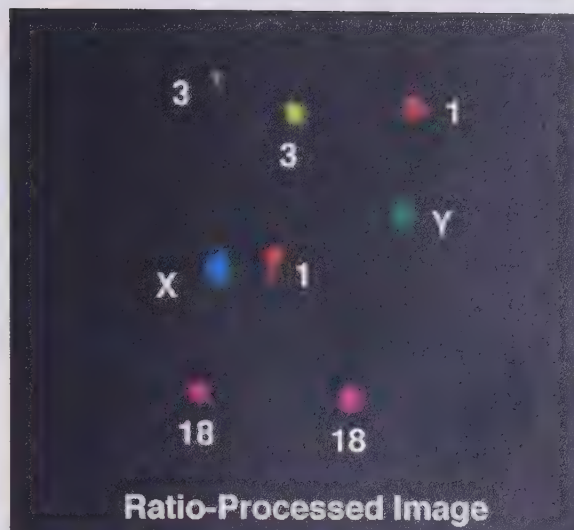


Plate 9 Multi-colour FISH by ratio-label processing in an interphase nucleus (see Chapter 13). The repeat-sequence probes were labelled as follows: chromosome 1, 100% biotin Texas Red; chromosome Y, 100% digoxigenin-FITC; chromosome 3, 50% biotin Texas Red, 50% digoxigenin-FITC; chromosome 18, 80% biotin-Texas Red,

20% digoxigenin-FITC; chromosome 3, 20% biotin-Texas Red, 80% digoxigenin-FITC. The image was acquired using a MRC 600 confocal microscope so that counterstaining of the nucleus was not possible. Original data provided by Yunling Zheng.



Plate 10 *Arabidopsis thaliana*. The picture shows a plant approximately 6 weeks old with numerous flowers and

fruits (siliques: arrow). The bar represents 1 cm. (see Chapter 34).

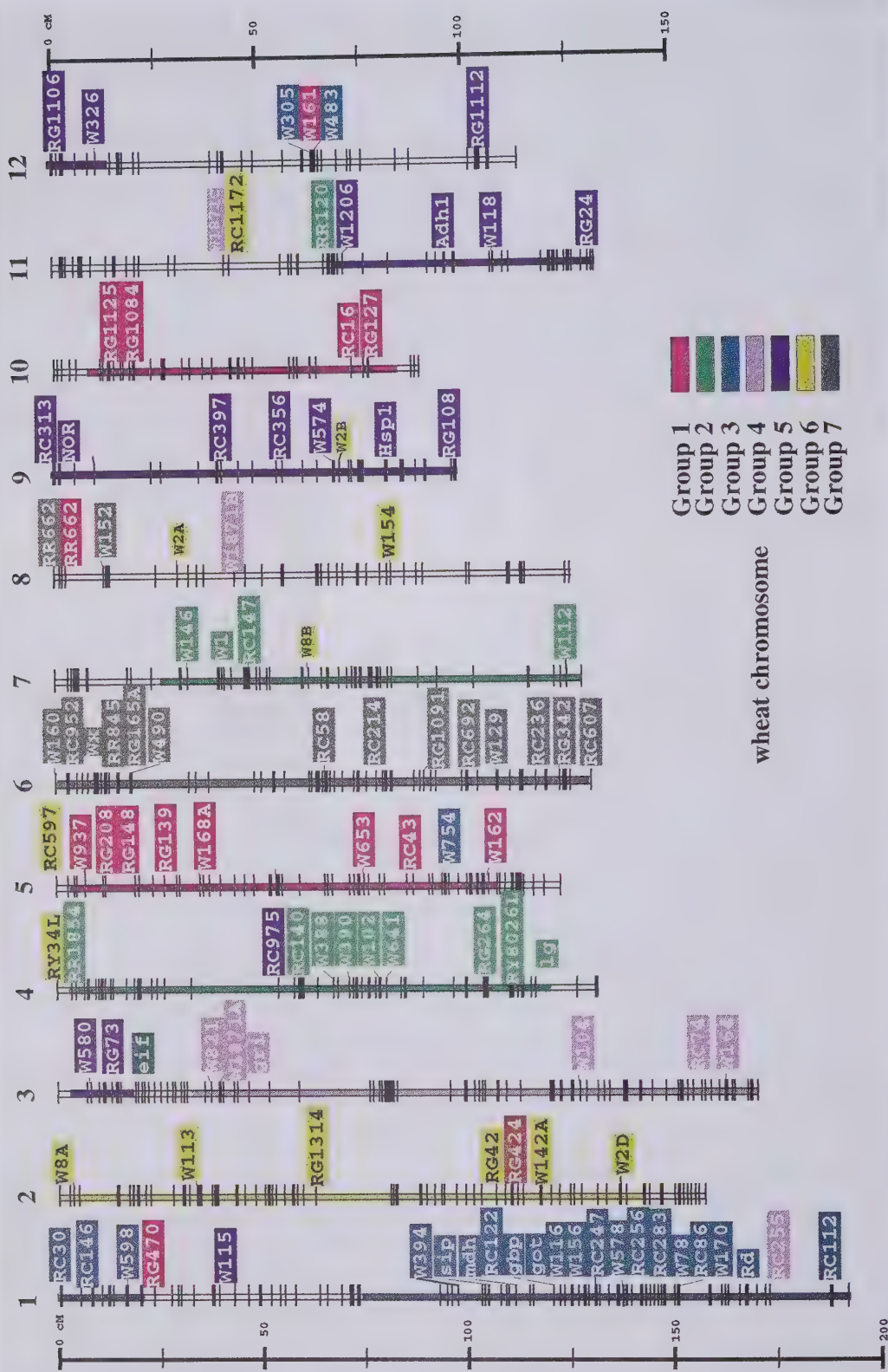


Plate 11 Synteny map between rice and wheat (see Chapter 34). On the rice RFLP map, the loci determined by wheat probes were indicated, with the loci of rice probes used for mapping on wheat RFLP map. The beginning letters indicate the sources of probes as follows: RC, rice cDNA clones from callus; RR, rice cDNA clones from root; RG, rice genomic DNA clones; RY, rice YAC clones; W, wheat DNA clones. Other abbreviations are as follows: sip, stress inducible

protein; mdh, malate dehydrogenase; gbp, GTP-binding regulatory protein β -chain; got, aspartate aminotransferase; Rd, red pericarp and seed coat; eif, initiation factor eIF-5A; gri, glycine-rich protein 2; lg, liguleless; Wx, waxy; NOR, nucleolus organizer region; hsp1, heat shock protein 1; adh1, alcohol dehydrogenase 1.

Chapter 11

Chromosome
microdissection

Debra M. Lillington & Andrew N. Shelling*

ICRF Department of Medical Oncology, Science Building, St Bartholomew's Hospital Medical College, Charterhouse Square, London EC1M 6BQ, UK

*Department of Obstetrics and Gynaecology, National Women's Hospital, Claude Road, Epsom, Auckland, New Zealand

11.1	Introduction, 265	11.8	Construction of the microclone library, 270
11.2	Chromosome preparation for microdissection, 265	11.8.1	Characterization of microclones, 271
11.2.1	Essential criteria, 265	11.8.1.1	Determination of insert size, 271
11.2.1.1	Availability and spreading of chromosomes, 265	11.8.1.2	Determination of the level of microclone redundancy, 272
11.2.1.2	Fidelity of DNA sequence, 265	11.8.1.3	Determination of the frequency of repetitive and unique sequence clones, 272
11.2.2	Sources of DNA damage, 265	11.8.1.4	Confirmation of the human origin and chromosomal region specificity, 272
11.2.2.1	Synchronizing agents, 265	11.9	Applications of microdissection, 273
11.2.2.2	Mitotic spindle inhibitors, 266		Protocol 55
11.2.2.3	Acid-induced damage, 266		Lymphocyte separation of peripheral blood samples, 274
11.2.3	Cell culture techniques for microdissection, 266		Protocol 56
11.2.4	Chromosome banding, 266		Cryopreservation of blood and bone marrow samples, 275
11.3	Microscopy and microdissection, 266		Protocol 57
11.3.1	Microscopy equipment, 267		Unsynchronized cultures, 276
11.3.2	Equipment and materials for preparing microneedles, 267		Protocol 58
11.3.3	Chromosome microdissection, 268		Harvesting chromosomes for microdissection and microcloning, 277
11.4	Microdissection and microcloning, 268		Troubleshooting, 278
11.4.1	Degenerate oligonucleotide primed PCR, 268		Poor chromosome spreading, 278
11.4.2	Universal DNA amplification procedure, 269		
11.5	Avoiding contamination, 269		
11.6	The microdissection amplification reaction, 270		
11.7	MicroFISH analysis, 270		

Protocol 59

Preparation of microneedles, 278

Troubleshooting, 279

Designing suitable needles, 279

Protocol 60

Chromosome microdissection, 279

Troubleshooting, 280

Recovering the microdissected chromosome fragment, 280

Protocol 61

DOP-PCR amplification reaction, 280

Troubleshooting, 281

No DNA staining in the gel, 281

DNA in the negative control lane, 281

Protocol 62

Fluorescence in situ hybridization using microdissected region-specific probes, 282

Troubleshooting, 285

No signal, 285

Lots of background signal, 285

Protocol 63

Microcloning of DOP-PCR amplification products, 286

Troubleshooting, 287

Low numbers of white colonies, 287

Too many blue colonies, 287

Protocol 64

Colony PCR to isolate microclone inserts, 287

Troubleshooting, 288

No band, 288

More than one band, 288

Smear of DNA, 289

Acknowledgements, 289

References, 289

11.1 Introduction

Since the first report of chromosome microdissection in 1981 using *Drosophila* polytene chromosomes [1], technical advances, particularly in the polymerase chain reaction (PCR) and DNA cloning methods, have established microdissection as a powerful tool in the analysis of the human genome.

In microdissection, a specific region of a chromosome is removed from the cell using microneedles designed specially for such microsurgery. The technique is attractive as it gives the operator direct access to any small region of a chromosome and enables them to remove it from the cell. The fragment of chromosome can then be analysed using a variety of molecular approaches.

Early experiments using microdissection and microcloning were time-consuming and cumbersome, requiring more than 100 fragments to be microdissected per investigation, and with relatively few microclones being isolated. Nowadays, microdissection and microcloning techniques can be applied to any source of chromosome, the chromosomes can be G-banded using trypsin and Giemsa (GTG-banded) for accurate identification, and, by incorporating PCR, the number of chromosome fragments which need to be microdissected is minimal (15–20 fragments is adequate). In fact, recent advances have allowed the amplification of DNA from only a single microdissected chromosome [2].

There are several approaches to microdissection which employ various pieces of equipment and microinstruments. This chapter will concentrate on

microdissection using an inverted microscope equipped with rotating stage and remote-controlled micromanipulator. However, the reader should bear in mind that other methods are being successfully used, such as laser microdissection [3] and microdissection using a microscope that incorporates an oil chamber in which the entire process of microdissection, DNA extraction, digestion with restriction enzymes and vector ligation is performed [4].

11.2 Chromosome preparation for microdissection

11.2.1 Essential criteria

11.2.1.1 Availability and spreading of chromosomes

A prerequisite for successful microdissection is a good supply of well-spread chromosomes. A chromosome that is well spread, with 'free space' either side of the target region, allows the operator room to manoeuvre the microneedle and therefore reduces the risk of contamination from neighbouring chromosomes. In addition, an abundance of chromosomes on the slide or coverslip saves much scanning time!

11.2.1.2 Fidelity of DNA sequence

Although techniques used in the cytogenetics laboratory to produce chromosome preparations are successful for routine cytogenetic analysis, they do cause extensive damage to the DNA. For microcloning and genome analysis, the integrity of the DNA is paramount and so steps must be taken to minimize the amount of DNA damage, as described below. If, however, microdissection is being undertaken solely to generate region-specific probes for microFISH (fluorescence *in situ* hybridization) analysis as described by Meltzer *et al.* [5], then conventional cytogenetic methods of chromosome preparation are suitable (see Chapters 7, 8 and 9 and references therein).

11.2.2 Sources of DNA damage

11.2.2.1 Synchronizing agents

Cell-synchronizing agents such as methotrexate, amethopterin, fluorodeoxyuridine and actinomycin D, which are used to obtain elongated chromosomes, are all potentially damaging to the DNA. Where the situation allows, unsynchronized cultures (see Protocol 57) are preferable and provide adequate numbers of good quality metaphases. When high-resolution (long) chromosomes are required, the use of thymidine to synchronize the cells is the recommended procedure (see Chapter 7,

Microdissection, PCR and microcloning or microFISH can be used to:

- construct region-specific genomic DNA libraries
- provide sequence-tagged sites (STS) for contig assembly and deletion mapping
- isolate region-specific cosmid and YAC clones
- generate region-specific polymorphic microsatellite probes
- identify the molecular structure and organization of specific chromosomal features such as centromeres, telomeres, satellites, G-light and G-dark bands
- assist cytogenetic analysis by identifying cryptic aberrations such as derivative chromosomes, ring chromosomes, markers, homogeneously staining regions (HSRs) and double minutes (dm)
- generate probes from regions involved in recurrent chromosomal rearrangements to aid the analysis of malignant disease

Protocol 56). Thymidine blocks the cells at a specific stage in the cell cycle; when the block is released the cells resume their cycling in synchrony. These cells are then harvested in early metaphase, yielding large numbers of divisions with long chromosomes. For microdissection, the chromosomes should be prepared on coverslips as described in Protocol 58.

11.2.2.2 Mitotic spindle inhibitors

Before chromosome harvest, cells need to be arrested at metaphase using a mitotic spindle inhibitor. Agents such as ethidium bromide should be avoided because although they enhance chromosome extension they cause nicking of the DNA. Colcemid is the standard agent used in most laboratories. A 10-min incubation with colcemid before harvest yields chromosomes which are in the early stages of metaphase and hence are more extended. Longer incubations with colcemid, up to 1 h, will result in a higher yield of metaphase spreads, but the chromosomes will be shorter as they become increasingly contracted towards the end of metaphase.

11.2.2.3 Acid-induced damage

Chromosomes are routinely 'fixed' in metaphase with methanol/acetic acid, 3:1, to remove cytoplasmic debris and maintain chromosome morphology. This fixation step is prolonged to give clean preparations of good quality for cytogenetic analysis. But acid causes depurination of DNA, and for the purpose of microdissection and microcloning this is undesirable. Adaptations to standard fixation procedures are therefore essential to limit the extent of damage (see Protocol 58). A prefix in 70% ethanol has been introduced as described by Kaiser *et al.* [6] followed by a short (10–20 s) fixation in methanol/acetic acid. This method of fixation is not suitable for whole blood samples as the length of fixation is insufficient to remove the debris from the red blood cells. For obtaining chromosomes for microdissection it is therefore advisable to remove the red blood cells before culturing (see Protocol 55) using a Ficoll gradient separation technique.

11.2.3 Cell culture techniques for microdissection

Essentially, microdissection can be performed on chromosomes obtained from any viable sample, including tumour tissue. Samples from which chromosomes are commonly derived include peripheral blood, bone marrow aspirates, lymph nodes, amniotic fluid, chorionic villus and lymphoblastoid cell lines. Most samples can be cultured using standard cytogenetic techniques [7] (see Chapters 7 and 8) but for reasons highlighted in the previous

section, it is advisable to remove the white blood cells from whole blood samples before culturing (see Protocol 55). Chromosomes can also be prepared from frozen samples of patient material which have been cryopreserved in liquid nitrogen using dimethyl sulphoxide (DMSO). Many laboratories routinely store excess patient material in liquid nitrogen tanks for use in future research. Protocol 56 gives a simple and effective method for such cryopreservation and includes a reliable technique for thawing and culturing cells from these samples. To maximize the chances of a successful culture which yields adequate numbers of metaphase chromosomes, cultures should be set up with the optimum cell density of 10^6 cells ml^{-1} . Various culture techniques are included in this chapter (Protocols 55–57) and a chromosome harvesting procedure (Protocol 58) that works well for microdissection is described.

11.2.4 Chromosome banding

Some chromosomes are easily recognizable in an unbanded preparation but, for most purposes, accurate identification of the chromosomal region of interest is only possible if the chromosomes are banded. It is better to band chromosomes, preferably under sterile conditions, on the same day that they are required as they are then softer and easier to cut. G-banding, which is the standard method of banding in most laboratories, can be achieved by treating the coverslips with trypsin and then staining in Leishman's stain (see Chapter 7, Protocol 62). For microdissection purposes, the saline solution and buffer are autoclaved and the Leishman's stain and trypsin solution is filtered through a millipore filter.

11.3 Microscopy and microdissection

Chromosomes that have been prepared and banded as previously described can now be microdissected. We shall concentrate here on one particular microdissection system—an inverted microscope fitted with micromanipulation equipment (Fig. 11.1)—which is convenient and reliable. Using glass needles which have been prepared in the laboratory (see Protocol 59), the chromosome regions are microdissected from coverslips (see Protocol 60). The needle tip with the chromosome fragment attached is then broken into an Eppendorf tube containing either PCR buffer or sterile water. When a sufficient number of chromosome fragments has been microdissected, they are amplified by PCR (see Protocol 61).

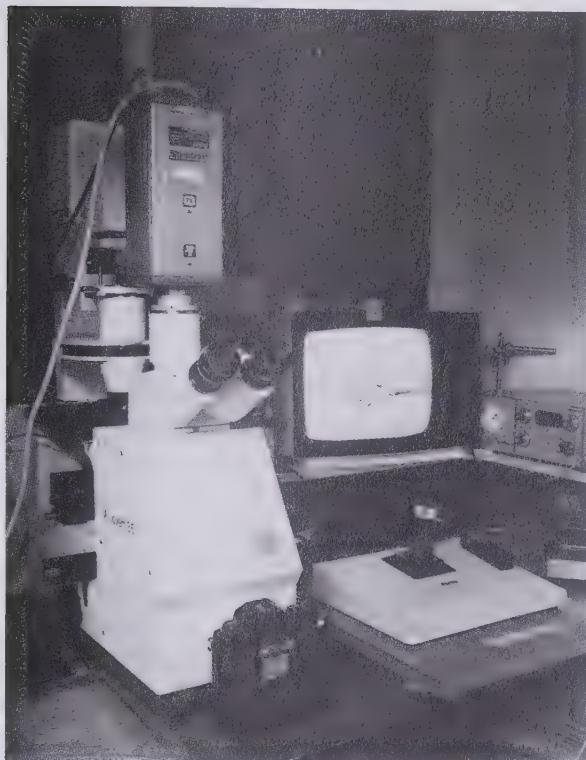


Fig. 11.1 Microscopy equipment for microdissection. Zeiss Axiovert microscope with gliding and rotating stage. Micromanipulation is carried out using the remote-controlled joystick. A video camera is attached to the microscope allowing microdissection to be visualized on the television screen. The glass needle can be seen on the screen to the right of the metaphase spread. From [24] by permission of Oxford University Press.

11.3.1 Microscopy equipment

Requirements:

- Zeiss Axiovert microscope;
- $\times 16$ eyepieces;
- $\times 4$ or $\times 6$ objective;
- $\times 63$ dry lens objective;
- gliding and rotating stage;
- mounts for micromanipulator;
- remote-controlled micromanipulator.

An inverted microscope such as Zeiss Axiovert (Fig. 11.1) mounted firmly on a vibration-free table is ideal for microdissection. The minimum attachments required include $\times 16$ eyepieces, $\times 4$ or $\times 6$ objective plus a $\times 63$ dry lens (or similar high power objective). Either brightfield or differential interference contrast optics are most suitable. To enable the target chromosome region to be correctly orientated for microdissection—that is, with the chromosome lined up vertically along the y -axis and the needle cutting horizontally through the x -axis—it is essential to be able to manoeuvre the coverslip in

all directions. This is achieved with a gliding and rotating stage in the centre of a fixed stage. The coverslip is placed over a central opening on the gliding and rotating stage with the objective underneath this central opening. Once the metaphase spread and target chromosome is selected, the stage can be rotated or dragged to suitably position the chromosome. The micromanipulator mounting points are attached to the fixed section of the stage. A long working distance condenser is preferable to allow adequate room for the micromanipulation. The micromanipulator has to be precisely controlled during operation and hence ones which require direct manual operation are less suitable. Remote-controlled electric or hydraulic micromanipulators are the best available and Zeiss manufacture an electrically controlled unit with extrasensitive movement control.

Various pieces of equipment can also be added to the above list, e.g. video camera, video printer and television screen, which are particularly useful for training purposes or for demonstration.

11.3.2 Equipment and materials for preparing microneedles

Requirements:

- 1- or 1.5-mm diameter glass rods or capillary tubes (borosilicate glass);
- microelectrode puller;
- grinder with lens system;
- microneedle holder;
- microoven.

A method for producing microneedles is described in Protocol 59. Both glass rod and thick-walled tubes are suitable. Before beginning microdissection you will need to select a suitable setting on the microelectrode puller for the type of glass you are using. A microelectrode puller with variable pulling force and heat setting (e.g. Campden Instruments microelectrode puller) allows you to control the length and shape of the needle tips. The needles are mounted into holders (Fig. 11.2) and the construction of these is such that the flattened sections of the stainless steel shaft when clamped into the grinder and subsequently the micromanipulator maintain the needle tip in the correct orientation. The needle itself is clamped into a miniature crocodile clip which has had the flat ends bent to clamp round the tubular needle. The needle is ground as described in Protocol 59 and such a grinder is available from Narashige. It is essential to remove any possible contaminants from the needle before use and hence at this stage the needle tip is heated to 300°C in a microoven. The

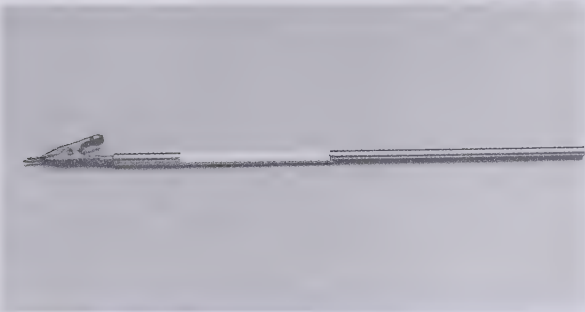


Fig. 11.2 Needle holder. The needle holder consists of a stainless steel shaft with flattened sections for clamping the holder into the grinder and micromanipulator. The glass needle is held in a miniature crocodile clip cemented to the end of the shaft.

microoven basically consists of a 5-mm glass tube surrounded by a heating element and insulating jacket. The temperature is controlled by a thermostat or electrothermal regulator.

11.3.3 Chromosome microdissection

Protocol 60 is designed for microdissecting a specific chromosomal region with the equipment described in the previous section. Figure 11.3 illustrates the microdissection of chromosome 6.



(a)

Fig. 11.3 Microdissection of region 6q26–27 from human chromosome 6. G-banded human metaphase chromosomes (a) before and (b) after microdissection of

11.4 Microdissection and microcloning

Microdissection and microcloning provide molecular geneticists with direct access to important chromosomal regions and enable them to isolate a large number of probes for analysis. Two of the main methods for constructing large human genomic libraries by microdissection and PCR-mediated microcloning methods are outlined below.

11.4.1 Degenerate oligonucleotide primed PCR

This method takes advantage of a primer that contains partially degenerate nucleotides to randomly amplify short fragments at frequently occurring priming sites within the genome. The degenerate oligonucleotide (DOP) primer (Fig. 11.4) and the principle of DOP-PCR was first described by Telenius *et al.* [8].

Briefly, priming occurs from the 3' ATGTGG nucleotides during the initial low annealing temperature cycles of the PCR reaction. These sequences occur frequently within the genome, at a similar frequency to restriction endonuclease recognition sites. The six degenerate nucleotides help to stabilize the specified 3' primer sequences,



(b)

the 6q26–27 region. The removed region is indicated by the arrow.

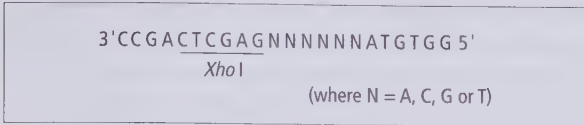


Fig. 11.4 Oligonucleotide primer for DOP-PCR.

by effectively allowing the primer to anneal as a 12mer. The 5' end of the primer contains the nucleotide recognition sequence for the *Xho*I restriction endonuclease, required for later cloning steps. These 5' sequences also allow primers to anneal efficiently to previously amplified DNA, thus allowing a higher annealing temperature during later PCR cycles.

The DOP-PCR method was initially devised as a rapid and efficient method of amplifying microdissected chromosomes, which did not require excessive technical expertise. It has worked successfully in a number of laboratories, and has been shown to give strong signals when amplified DNA is hybridized back to normal metaphase chromosomes by FISH. Recently the method has been fine-tuned to allow amplification from a single microdissected chromosome [2]. One major disadvantage of DOP-PCR, however, is the high risk of contamination (as with most PCR amplifications from small numbers of DNA fragments). Additionally, uneven hybridization to metaphase chromosomes has been reported [9,10], where DOP-PCR amplified microdissected DNA often fails to paint repetitive sequences in acrocentric short arms, at the centromere, at the telomeres, and in some heterochromatic regions.

11.4.2 Universal DNA amplification procedure

The second group of methods make use of restriction endonuclease digestion and DNA ligation steps directly on the microdissected material before amplification. As there is a minimal amount of DNA, microchemical techniques are performed on a nanolitre microdrop contained in an oil chamber. There are two main variations on a similar overall procedure. The first was initially used by Lüdecke *et al.* [11] in which the microdissected DNA was digested with the blunt-end restriction endonuclease, *Rsa*I. These fragments were then cloned into a *Sma*I-cut pUC13 vector and amplified by PCR using the plasmid vector sequencing primers. The amplified DNA inserts are cleaved with a second restriction enzyme (*Eco*RI) that flanks the cloning site and are subcloned into a second plasmid vector to generate the microclone library.

A second similar method, known as the linker-adaptor PCR (LA-PCR) method, was devised by Saunders *et al.* [12] and Johnson [13]. This method ligates microdissected DNA to linker adaptors before PCR amplification rather than to a plasmid vector. The dissected DNA is digested with a frequent cutting restriction endonuclease (such as *Sau*3AI or *Mbo*I), ligated to a 5' protruding *Mbo*I linker adaptor consisting of phosphorylated 24mer and dephosphorylated 20mer oligonucleotides, and amplified using the 20mer DNA as a primer. The PCR products are then digested with *Mbo*I to remove the adaptor, and ligated into the *Bam*HI site of a suitable plasmid vector.

Several libraries have been constructed using both these universal DNA amplification PCR methods [14], and typically contain large numbers of microclones. However, these methods are technically difficult, and involve working with small quantities of DNA in nanolitre microdrops contained in an oil drop.

11.5 Avoiding contamination

The process of PCR amplification allows the generation of microgram quantities of DNA from only a few microdissected chromosomes (femtogram quantities). This represents an amplification of approximately 10⁹-fold. Any contaminating DNA molecule, whether airborne from the laboratory (plasmid, bacteria, phage, yeast), or from material which has previously been amplified by PCR will also be amplified. Therefore stringent procedures must be followed to minimize contamination during every step in the microdissection/microcloning protocol. Along with the basic precautions normal for PCR [15], the following procedures are useful to reduce contamination [14].

- Use only sterile disposable pipettes, flasks, tubes, etc.
- Wear gloves at all stages and change frequently.
- Prepare buffers and reagents using aseptic techniques and place in sterile tubes for single use only.
- Expose micropipettes, buffers and reagents to UV (254 nm).
- Autoclave all buffers and reagents.
- Use micropipette for one operation only.
- Include several control reactions.
- Physically separate pre- and post-PCR manipulations.
- Minimize the number of manipulations.

A certain level of contamination is inevitable and it is a matter of reducing it to manageable levels. Low-level contamination is not in practice a serious problem as the clones containing human inserts can

be selected by hybridization to a human genomic Southern blot [15]. Virtually all microclones that have been shown to be of human origin have been derived from the dissected region, presumably as a result of the high precision of the microdissection technique [14].

11.6 The microdissection amplification reaction

Before attempting to amplify the microdissected DNA, it is advisable to perform some trial reactions to test that the PCR protocol is working efficiently, for example by amplifying a range of dilutions of genomic DNA (100 ng to 10 fg).

Various control tubes are also essential in the PCR reaction alongside the tube containing the microdissected DNA. A tube containing blank needle tips and a tube containing only the 10× *Taq* polymerase collection buffer are recommended as controls since neither of these contain any template DNA.

Protocol 61 describes the DOP-PCR amplification for microdissected chromosome fragments, but other methods can be found elsewhere [11, 16].

Figure 11.5 shows amplification products of a DOP-PCR reaction. Amplification can be seen in lanes 2 and 3 which contain microdissected DNA from chromosome 6q26–27 region. The two negative control lanes (5 and 6) are blank. The 6q26–27 lanes contain a smear of DNA from about 200–1000 bp and also some distinct bands. The source of these over-represented bands is unknown. They possibly represent contamination or could be the result of preferential amplification of a sequence within the

microdissected region. This preferential amplification may be due to the identity of the DOP-PCR primer with subtelomeric or telomeric repetitive sequences. The microFISH data (Plate 7) offers some support for this hypothesis.

11.7 MicroFISH analysis

Following DOP-PCR amplification of the microdissected chromosomal DNA, microFISH [5] should be performed to check the integrity of the DNA before microcloning is undertaken. MicroFISH is a reverse painting technique which can be used to elucidate the origin of the template DNA.

Briefly, an aliquot of DNA from the PCR reaction is purified by a purification column (Promega Wizard PCR preps column). One microgram DNA is then labelled with biotin-11-dUTP by nick translation (see Chapter 9, Protocol 44) and purified using a sephadex G-50 column. MicroFISH is then carried out as described in Protocol 62.

Plate 7 shows a microFISH result from the chromosomal region 6q26–27.

11.8 Construction of the microclone library

Having confirmed by microFISH that the amplification products are derived from the chromosome region originally microdissected, microcloning can now be performed. The amplification products must be digested and ligated into an appropriate site in a plasmid vector. For the DOP-PCR protocol, a *Xho*I restriction site is present at each end of the amplified

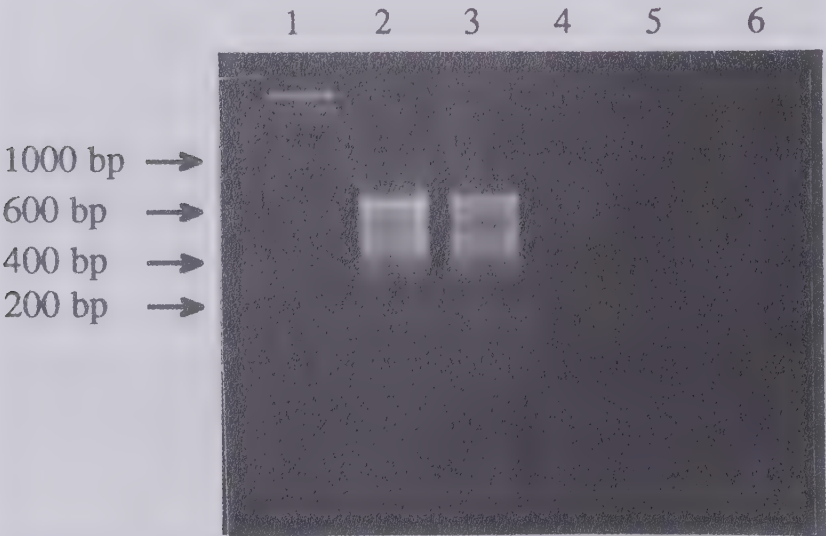


Fig. 11.5 A 2% gel showing the general size distribution of the amplified microdissected 6q26–27 region. Five microlitres from each amplification reaction separated on a 2% agarose gel. Lanes 2 and 3 show DNA amplified from the microdissected region. Lanes 4 and 5 are control lanes, containing the microdissection glass tips, and collection buffer with no target DNA, respectively. Lane 1 is the marker lane containing λ DNA digested with *Hind*III and *Eco*RI.

fragments. The procedure for microcloning such fragments is described in Protocol 63. The efficiency of the ligation reaction and the efficiency of the transformation step will determine the number of clones realized. For most purposes, only a few hundred colonies need to be analysed in detail, although in many situations several thousand colonies may be needed (e.g. to screen a large microclone library for the presence of microsatellite sequences (see Chapter 5). For most purposes, 100 ng PCR product ligated with 400 ng plasmid vector (Bluescript; Stratagene) in a 10- μ l volume ligation at 14 °C overnight, followed by transfection of high-efficiency competent cells (DH5 α), should give more than 10 000 colonies.

The number of clones to be analysed often depends on what the microclone library is going to be used for. To create a yeast artificial chromosome (YAC) contig from a single chromosomal band (assuming 10 000 kilobases (kb) per band), a library of 200 clones will give five clones every 500 kb (the average size of a YAC molecule), which should be sufficient to generate a YAC contig. To screen for microsatellites, 20 000 colonies may need to be analysed.

11.8.1 Characterization of microclones

Once the microclone library has been constructed, it should be tested to verify that it contains a number

of unique clones that are derived from the microdissected region. The following steps are important guidelines for characterizing the library.

11.8.1.1 Determination of insert size

Inserts can be quickly recovered by PCR amplification of individual colonies using vector primers such as the M13 forward and reverse sequencing primers (Protocol 64). Each cloned insert should be sized on a 2% agarose electrophoresis gel (Fig. 11.6). The analysis of the inserts will provide a great deal of information about the library. Typical libraries should contain a range of sizes (about 100–2000 base pairs) which should reflect the size distribution of the original amplification reaction. The average size of microclone inserts is usually about 400 base pairs. Smaller sized insert fragments have been observed in some microclone libraries, presumably due to the preferential amplification and/or cloning of small inserts [11], and there is also some evidence that hydrolysis of DNA by acid treatment during fixation of the chromosome may cause smaller than expected inserts [17]. Clones that do not contain an insert will show a single band corresponding to the distance between vector primers in pBluescript. Occasionally, some clones do not amplify at all and should be repeated (see Troubleshooting, Protocol 64). We have found on rechecking nonamplified clones that they do often contain larger inserts.

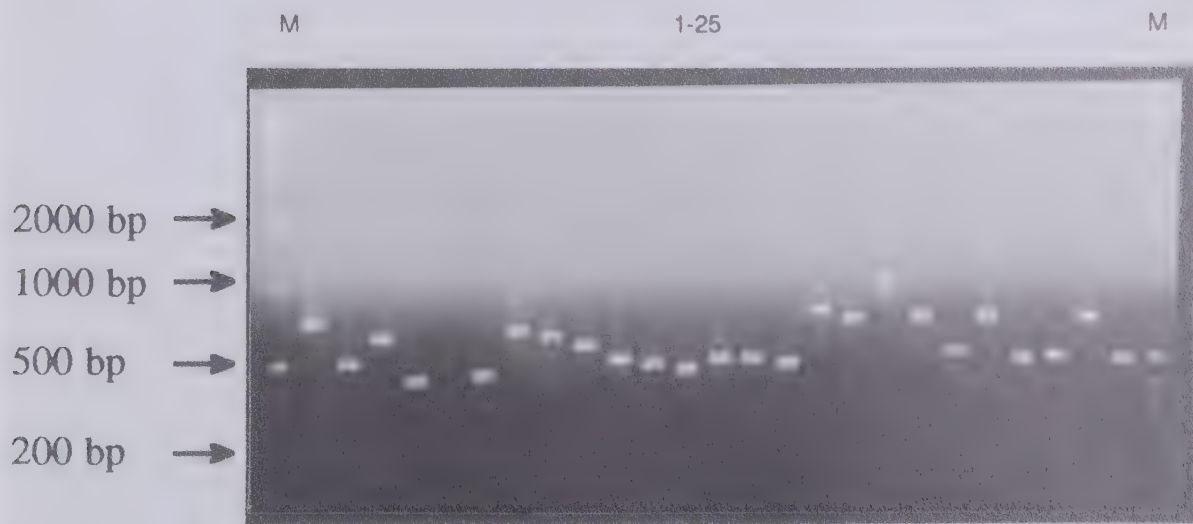


Fig. 11.6 A 2% gel showing the size of the amplified inserts of 25 microclones sampled from the chromosome 6q26–27 library. Ten microlitres from each of the

amplified inserts has been separated on a 2% agarose gel (lanes 2–26). Lanes labelled M contain marker DNA (1 kb DNA ladder; Gibco-BRL).

11.8.1.2 Determination of the level of microclone redundancy

It is quite useful to determine the microclone redundancy—that is, the number of over-represented clones. This is easily checked by labelling a suspected redundant clone using the random priming method of Feinberg and Vogelstein [18] and probing a Southern blot of the entire microclone library as follows:

- 1 Separate the inserts in 2% gel.
- 2 Transfer to nylon membrane (Hybond N⁺) under alkaline conditions.
- 3 Prehybridize at 65 °C for 2 h [19].
- 4 Add denatured probe and hybridize for 16–20 h at 65 °C.
- 5 Wash in 0.1×SSC/1% SDS at 65 °C for 30 min.
- 6 Expose membrane to X-ray film at –70 °C with DuPont intensifying screens.

In a library that we have constructed, two over-represented families were identified by Southern blot, one of which made up 20% of the library, and the other 5.5%. It is unclear why these families have occurred; they may either be contaminants or the result of amplification of a weakly repetitive sequence which contains the recognition site of the amplification primer.

11.8.1.3 Determination of the frequency of repetitive and unique sequence clones

Southern blots of the microclone inserts are hybridized

with ³²P-labelled total human DNA (Fig. 11.7). Microclones with highly repetitive sequences (giving very strong hybridizing signals), middle and low repetitive sequences (intermediate or weak signals) and unique or very-low-copy repetitive sequences (with no hybridizing signals) can easily be identified. We have also used an Alu-probe, which hybridized to all the highly repetitive clones, and a human LINE probe, which did not hybridize to any of the 400 microclones tested.

11.8.1.4 Confirmation of the human origin and chromosomal region specificity

A number of unique-sequence microclones are chosen, labelled to high specific activity with ³²P, and hybridized to a Southern blot containing digested human DNA, somatic cell hybrid DNA containing the chromosomal region of interest, DNA from the background species of the somatic cell hybrids, and perhaps a number of other species (zoo blot) to identify species-conserved sequences. The proportion of clones that hybridize to human and hybrid DNAs provides a good indication of the quality of the microclone library. Short probes such as microclone probes do not make good hybridization probes, often giving weak signals. To overcome these problems, 10 µg genomic DNA should be digested, and only allowed to separate a short distance on an agarose gel. In addition, higher specific activity probes can be made (> 10⁸ d.p.m. µg⁻¹), and

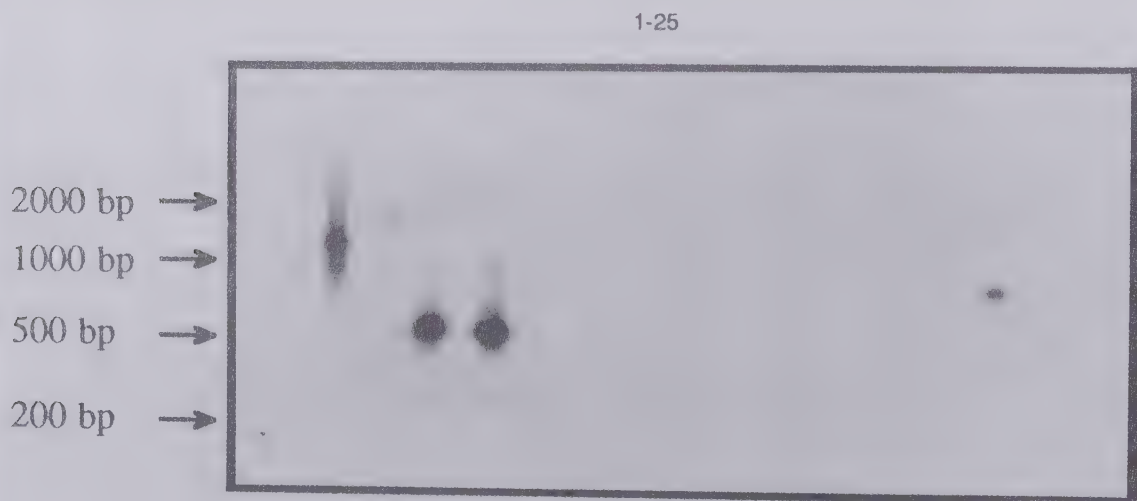


Fig. 11.7 Southern blot of amplified inserts from the chromosome 6q26–27 library microclones hybridized with ³²P-labelled total human genomic DNA. The

amplified inserts separated on a 2% agarose gel (Fig. 11.6) were transferred to a nylon membrane and hybridized with ³²P-labelled total human genomic DNA.

the washing stringency can be lowered in the final wash to $0.5\times\text{SSC}/1\%$ SDS at 55°C . Some groups use a PCR-labelling technique to ensure that they gain full-length probes from their short microclone inserts, e.g. [20], rather than shorter probes generated by the random priming method of Feinberg and Vogelstein [18].

11.9 Applications of microdissection

Microdissection and microcloning, in conjunction with PCR, enables the construction of region- or band-specific genomic libraries which will aid in building high resolution maps for the identification of disease-related genes within genomic regions [14]. DNA sequence markers are commonly generated from chromosome-specific genomic DNA libraries derived from somatic cell hybrids, radiation-induced hybrids or flow-sorted chromosomes using a variety of techniques. Microdissection and microcloning offers a refinement to other methods, enabling much smaller chromosome regions to be studied.

The microcloned DNA can be used for contig assembly and high resolution mapping. Pooled microclones are used to probe and isolate genomic libraries containing larger DNA inserts such as cosmids or YACs. In addition, each microclone can be sequenced and thus becomes a sequence-tagged site (STS) [21] which can provide physical landmarks within the microdissected region. These STSs can then directly aid the assembly of YACs into an overlapping contig.

Pools of microclones can also be hybridized to cDNA libraries to isolate expressed genes as potential candidate genes in disease.

Region-specific polymorphic microsatellite probes (see Chapter 5) can also be generated by microdissection using microsatellite primer probes to probe the microclone library. These can subsequently be converted to genetic markers for use in loss of heterozygosity or linkage analysis studies.

Any region of a chromosome can be microdissected, such as centromeres, telomeres, G-light bands, G-dark bands and satellites, and by analysing microclones from these regions the molecular structure and organization of these genomic landmarks can be identified.

Microdissection and microFISH is a valuable aid in cytogenetic analysis, permitting the characterization of many unresolved cytogenetic aberrations. The origin of cryptic translocations, ring chromosomes, derivative chromosomes, markers, homogeneously staining regions and double minutes can be elucidated by microFISH.

Microdissection has already been successfully employed to identify and generate probes from regions involved in chromosomal rearrangement and in deletions, and this approach will assist in the identification of novel genes associated with various diseases (see Case Study). The generation of probes from translocation breakpoints such as the *bcr-abl* junction [22] will be valuable in the analysis of malignant cells. The microdissection of breakpoint regions was first described by Cotter *et al.* [23] where, using gene-specific primers for PCR amplification, translocation breakpoints in malignant disease were mapped relative to known genes.

The isolation of genes underlying inherited and acquired genetic diseases will hopefully aid in the diagnosis, prevention and therapy of these diseases. Several disease loci have already been analysed by microdissection and microcloning and this microtechnology should continue to be a powerful tool in achieving the goals of future research projects.

Isolation of expressed sequences encoded by the human Xq terminal portion using microclone probes generated by laser microdissection.

Several region-specific microdissection libraries have been constructed recently and used for a variety of applications. Yokoi *et al.* [26] have shown how the construction of a region-specific library led to the isolation of candidate disease genes. The region they chose to investigate was the distal portion of chromosome Xq, which is known to house a variety of genes for neurological and neuromuscular disorders (see Appendix VII). Some of these genes have been identified but many have yet to be isolated. Using laser microdissection, Yokoi *et al.* microdissected the distal region of two homologues of chromosome Xq, amplified the DNA using a single unique primer, and constructed a regional genomic library containing 2×10^7 clones with an average insert size of 234 base pairs. Thirty per cent of the microclones contained unique sequence and 56% of these mapped to chromosome X. Expressed sequences were isolated by screening human brain cDNA libraries with pools of clones from the genomic microdissection library and 28 unique cDNA clones were detected in this way. Ten of the cDNA clones were shown to be nonoverlapping and each mapped to chromosome Xq. The 10 cDNA clones were completely sequenced and no significant homology to previously characterized primate genes was found. One clone in particular, which mapped to Xq27.3-qter and contained an open reading frame of 281 amino acids, was judged to be a probable coding sequence and was shown to be expressed in all eight tissues tested. This clone may be a candidate gene for one of the 19 or more unidentified disease genes which map to this region on chromosome X. Further studies are in progress to determine whether any of these isolated sequences represent novel candidate genes for these X-linked heritable disorders.

Case Study 11.1

Protocol 55 Lymphocyte separation of peripheral blood samples

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.

Materials

- lymphocyte separation medium (e.g. 'Lymphoprep' or Ficoll–Hypaque) (Nycomed)
- serum-free RPMI 1640 medium (Gibco-BRL)
- complete medium: 100 ml RPMI 1640 (Gibco-BRL) with Glutamax-1^a, 20 ml fetal calf serum (FCS), 1 ml penicillin/streptomycin (P/S)

Method

Use between 10 and 20 ml of fresh heparinized peripheral blood.

- 1 Dilute fresh peripheral blood with an equal volume of serum-free medium.
- 2 Place 3 ml of lymphocyte separation medium into a centrifuge tube and carefully layer 7 ml of diluted blood on top by running it down the side of the tube with a plastic pipette.
- 3 Centrifuge at 1500 r.p.m. for 20 min. The red cells will pellet at the base of the tube and the white cells will form a buff-coloured layer between the separation medium and the blood serum.
- 4 Collect the buffy coat layer using a 1-ml syringe while avoiding drawing up the separation medium.
- 5 Place the white blood cells into a fresh centrifuge tube^b and top up to the 10-ml mark with serum-free medium.
- 6 Centrifuge at 1000 r.p.m. for 10 min.
- 7 Remove supernatant and resuspend cells in a further 10 ml of serum-free medium.
- 8 Centrifuge at 1000 r.p.m. for 10 min.
- 9 Remove supernatant and add approximately 2 ml of culture medium.
- 10 Assess white cell count.^c
- 11 Set up cultures as described in Protocol 57.

^a Glutamax-1 is L-alanyl-L-glutamine.
^b Pool white cells from same blood sample at this stage if several tubes were initially used.
^c White blood cell count can be measured using a coulter counter or alternatively using a haemocytometer.

.....

Protocol 56 Cryopreservation of blood and bone marrow samples

Overview

- (a)** Freezing cells viably
- (b)** Thawing cryopreserved cells

(a) Freezing cells viably

Materials

- freezing medium: Gibco RPMI 1640 medium (2 parts), FCS (2 parts), DMSO (1 part)

Method

It is best to freeze blood or bone marrow samples which have been freshly separated as in Protocol 55.

- 1** Assess the white blood cell count and resuspend in serum-free medium at a concentration of between 2 and 10×10^7 cells ml⁻¹.
- 2** Add 0.5 ml of DMSO freezing medium to 0.5 ml of cells per vial, dropwise with constant agitation.^a
- 3** Cool or freeze vial immediately.^b Introduce vial into liquid nitrogen gradually.^c

(b) Thawing cryopreserved cells

Materials

- thawing medium: 27 ml RPMI 1640, 3 ml FCS
- complete medium (see Protocol 55)

Method

- 1** Remove vial from liquid nitrogen and thaw quickly in a 37 °C waterbath.
- 2** Place cells into a 10-ml centrifuge tube.
- 3** Add 5 ml thawing medium, one drop every 10 s for 2 min, then two drops every 10 s for 2 min and then gradually increase the number of drops until 5 ml has been added.
- 4** Add a further 5 ml thawing medium.
- 5** Centrifuge at 2000 r.p.m. for 5 min.
- 6** Remove supernatant and repeat steps 3–5 twice.

^aThe final concentration of cells is 1–5 x 10⁷ in 10% DMSO per vial.

^bDMSO is toxic to cells at concentration above 1% at room temperature. Therefore cool vial immediately to minimize this effect.

^cCryopreservation bins usually incorporate a mechanism that allows gradual introduction of the vial into the liquid nitrogen. Alternatively, steps of gradual cooling at temperatures between 4 °C and –70 °C can be used.

- 7 Finally resuspend cells in 2 ml medium and perform a white blood cell count.
- 8 Set up cultures containing 10^6 cells ml⁻¹ in complete medium for between 48 and 96 h.

.....

Protocol 57 Unsynchronized cultures

Overview

- (a) Constitutional blood culture
- (b) Leukaemic blood or bone marrow culture
- (c) Other sample types (e.g. amniotic fluid, chorionic villus, cell lines and solid tumours)

Materials

- complete medium (see Protocol 55)
- phytohaemagglutinin (PHA)
- colcemid stock solution (10 µg ml⁻¹)

(a) Constitutional blood culture

Method

Use separated lymphocytes or whole peripheral blood.^a

- 1 Place 0.4 ml fresh heparinized blood or appropriate volume of separated lymphocytes (10^6 cells ml⁻¹) in 10 ml complete medium.
- 2 Add 0.2 ml PHA.
- 3 Incubate at 37 °C for 72 h.

- 4 Add 0.1 ml colcemid (final concentration 0.01 µg ml⁻¹) for between 10 and 60 min.^b

(b) Leukaemic blood or bone marrow culture

Method

Set up 5 ml cultures in complete medium containing 10^6 cells ml⁻¹, incubate at 37 °C, and harvest either:

- after 24 h incubation with 0.05 ml colcemid added for the last hour; or
- after overnight incubation with 0.025 ml colcemid.

^aFor detailed molecular analysis of the microdissected DNA, the fidelity of the sequence is important. Separated lymphocytes and short fixation time are therefore necessary. Whole blood cultures and standard fixation methods are suitable for generating microFISH probes.

^bThe length of incubation with colcemid will depend on the length of chromosome and the yield of mitoses required.

- (c) Other sample types (e.g. amniotic fluid, chorionic villus, cell lines and solid tumours)

Method

- 1 Set up using standard cytogenetic procedures [7].
- 2 When there are sufficient numbers of cells present, add colcemid (final concentration, 0.01 µg ml⁻¹) for 10–60 min.

.....

Protocol 58 Harvesting chromosomes for microdissection and microcloning

Materials

- colcemid stock solution (10 µg ml⁻¹)
- 0.075 M KCl
- prefix (70% ethanol)
- fixative: 3 parts methanol, 1 part glacial acetic acid

Method

To be performed in a sterile safety cabinet using aseptic techniques.

- 1 Add 0.05 ml colcemid to 5 ml cultures for 10–60 min.
- 2 Centrifuge at 1000 r.p.m. for 10 min.
- 3 Remove supernatant and add an equal volume of KCl (prewarmed at 37 °C) for 10–15 min.
- 4 Centrifuge at 1000 r.p.m. for 10 min.
- 5 Remove supernatant and add 5–10 ml prefix.
- 6 Place tube at –20 °C for at least 30 min.
- 7 Centrifuge at 1000 r.p.m. for 10 min.
- 8 Remove all but approx. 0.25 ml of supernatant and resuspend pellet.
- 9 Add between 0.5 and 2 ml (depending on pellet size) fresh ice-cold fixative within 20 s while constantly agitating tube.
- 10 Immediately drop onto clean^a ice-cold coverslips^b then place a drop of prefix on top.
- 11 Allow to dry in sterile hood then store at –20 °C in sterile container.

^a Clean coverslips in methanol and place in sterile container in freezer to achieve better spreading.
^b Coverslips are preferentially used because glass slides are too thick for the objectives on the inverted microscope system.

.....

Troubleshooting

Poor chromosome spreading

Proper fixation is recognized as probably the most critical factor in acquiring well-spread chromosomes. This protocol, however, employs an extremely short fixation step and this can lead to difficulties in spreading. Other factors, such as relative humidity and the ambient temperature, also have an effect. Although there are no easy solutions to the problem of underspread chromosomes, one or a combination of the following may be helpful:

- *Spread onto ice-cold coverslips.*
- *Spread onto warmed coverslips either on a moist tissue on a hotplate or on a rack in a water bath.*
- *Use wet coverslips soaked in methanol or sterile distilled water.*
- *Shortly after dropping the cell suspension onto the coverslip, place a drop of fixative or 70% ethanol on top.*
- *Alter the angle or height at which the suspension is dropped onto the coverslip.*

Overspread chromosomes are in fact preferable for microdissection purposes, allowing easier access to the region of interest.

Protocol 59 Preparation of microneedles

Materials

- 1.5-mm diameter glass rod or glass tube

Method

- 1 Mount a 7.5-cm length of glass rod or capillary tube into the microelectrode puller.
- 2 Pull glass slowly to produce two short needles with a fine point.
- 3 Mount the needles into the microneedle holders (Fig. 11.2) and clamp holder into grinder at an angle of 40°.
- 4 While viewing needle tip through the lens system and applying drops of water to the grinding wheel, lower the needle until the tip is just touching the wheel.^a
- 5 Wash the needle in a stream of 70% ethanol.
- 6 Carefully place the needle in the microoven set at 300 °C for 30 s.
- 7 Allow needle to cool before using.

^aIf glass rod is used to prepare the needle, grind tip until a faint bright spot appears. If glass tube is used, grind tip until the water just begins to rise up the tube.

Troubleshooting

Designing suitable needles

Microneedles need to be strong enough to cut through the chromosomal material without breaking but fine enough to cut a narrow region. Many different needles may need to be tested before the ideal design is reached. Once a suitable programme on the microelectrode needle puller has been identified, this should produce consistently good needles for the type of glass being used. A combination of the following will alter the shape of the needle tips being produced on the microelectrode puller:

- *adjust variable pulling force;*
- *decrease or increase the temperature.*

If the needle tip is too fine, grinding for longer periods may be effective. If the needle tip is too stubby, minimal grinding may be the answer.

Broken needles are usually a result of carelessness in transferring from one piece of equipment to another and hence extra care should be taken.

Occasionally a batch of glass may be at fault and if several broken needles are seen it may be worth using a different type of glass. A fine needle tip may explode in the microoven if the temperature is too extreme.

Protocol 60 Chromosome microdissection

Method

- 1 Locate an easily accessible target chromosome and centralize it in the field of view. Rotate stage until chromosome is perpendicular relative to the direction of the needle.
- 2 Place needle holder in micromanipulator and with $\times 6$ objective and condenser diaphragm partly closed, manually move and lower the needle towards the coverslip until the tip of the needle is just above the metaphase.
- 3 Change to a high-power dry lens and close condenser diaphragm further.^a Lower the needle until the tip is just visible but not in contact with the coverslip using the remote-controlled joystick for accurate control.
- 4 Open condenser to view chromosomes^b and lower the needle tip further using fine movement controls until the tip appears in the field of view.

^a Closing the condenser diaphragm creates an apparent depth of field so the needle will be in focus well above the coverslip.

^b On opening the condenser the needle tip will not be in the field of view until it is lowered further towards the coverslip.

- 5 Move the needle to the edge of the target chromosome region.
- 6 Lower further until the needle tip just touches the coverslip.
 - (a) Slowly lowering the needle further will cause the needle to move forward through the chromosome under its own weight.
 - (b) Alternatively, the needle can be moved manually through the chromosome using joystick controls.
- 7 As the needle cuts through the chromosome, the fragment folds up onto the top side of the needle tip.
- 8 Lift the needle away from the chromosome and coverslip and break needle tip directly into Eppendorf tube containing 10× *Taq* polymerase buffer using pressure against the side of the tube.

.....

Troubleshooting

Recovering the microdissected chromosome fragment

If the fragment does not automatically adhere to the needle tip but is merely pushed out to the side of the chromosome then it can usually be picked up by gentle prodding with the needle tip. Chromosomes that have been banded on the same day, and are therefore damp, tend to stick to the needle better than dry chromosomes. The size and shape of the needle tip will also affect the ability of the needle to pick up the fragment and you will need to experiment with the design before starting on your precious material. If you need to microdissect a very narrow region you may find it easier to microdissect the two areas adjacent to the region of interest first with one needle and discard these, and then pick up the piece in the middle with a second needle. The microdissected fragments should be visible on the tip of the needle under the microscope. It is not advisable to chase fragments round the cell in an effort to pick them up as this increases the chance of contamination.

.....

Protocol 61 DOP-PCR amplification reaction

(Modified from Guan *et al.* [2])

Materials

- reaction mixture: 5 µl 10× *Taq* polymerase buffer, 2 µl MgCl₂ (50 mM stock), 2 µl dNTPs (5 mM stock), 4 µl DOP-PCR primers (20 µM stock), 0.5 µl *Taq* polymerase (5 U µl⁻¹ stock), 36.5 µl sterile H₂O, mineral oil

Method

- 1 Collect microdissected chromosomal fragments in 5 µl 10× *Taq* polymerase buffer.^a
- 2 Add remainder of reaction components listed above (50 µl total volume).
- 3 Overlay with a drop of mineral oil and amplify using the following thermal cycle programme:
94 °C for 10 min;
8 cycles of:
94 °C for 1 min;
30 °C for 1 min ;
28 cycles of:
72 °C for 3 min;
94 °C for 1 min;
56 °C for 1 min ;
72 °C for 3 min;
and then:
72 °C for 10 min.
- 4 Remove 1 µl amplification product from first round PCR^c above and place into a new reaction.
- 5 Repeat last 28 cycles of above programme with a final 10-min extension period at 72 °C.
- 6 Assess PCR products by separation on agarose gel (Fig. 11.5).

^aSome protocols suggest using a proteinase K and SDS step followed by phenol extraction. We and others have found that omitting this step has no apparent effect on amplification or subsequent probe quality.

^bAdd *Taq* polymerase to the reaction at the end of the 94 °C denaturing step (Hot Start PCR) to minimize inactivation of the *Taq* polymerase enzyme and to prevent nonspecific primer extension during the precycling period.

^cOne round of amplification is usually insufficient to achieve detectable amounts of DNA.

Troubleshooting

No DNA staining in the gel

As a first step it is important that you confirm that you are able to amplify DNA from a range of dilutions of total genomic DNA. Try a range of concentrations such as 100 ng to 10 fg. This will confirm that your primers and PCR conditions are satisfactory. Other things to try are:

- increasing the number of chromosome fragments;
- increasing the primer concentration;
- increasing the number of PCR cycles.

DNA in the negative control lane

Throw away all buffers and enzymes and start again. Repeat using sterile conditions as outlined in the text. If the microdissected material is contaminated, then the microdissection must be repeated.

Protocol 62 Fluorescence in situ hybridization using microdissected region-specific probes

Overview

- (a) Competition step
- (b) Slide preparation
- (c) Denaturation of chromosomal DNA
- (d) Hybridization
- (e) Posthybridization washes
- (f) Detection of biotin label

General reagents

- 20×SSC (pH 5.3)
- 2×SSC (pH 7)
- 4×SSCT: 4×SSC, 0.05% Triton X-100, pH 7
- PBS (pH 7)
- formamide
- ethanol series (70%, 95%, 100%)

(a) Competition step

Materials

- Cot-1 DNA (Gibco-BRL)
- biotin-labelled probe^a
- 3 M sodium acetate
- ethanol absolute
- hybridization buffer: 50% formamide, 10% dextran sulphate, 2×SSC, 1% Triton X-100, sterile distilled water

Method

- 1 Place 200 ng labelled probe into 1.5-ml tube on ice.
- 2 Add 5 µl Cot-1 DNA.
- 3 Add $\frac{1}{10}$ volume 3 M sodium acetate and 2 vols ice-cold ethanol.
- 4 Place tube at −20 °C overnight or at −70 °C for 1 h.
- 5 Microfuge tube for 20 min.
- 6 Pour off supernatant and invert tube on tissue until pellet is dry.
- 7 Add 15 µl hybridization buffer to the pellet and mix gently by pipetting.
- 8 Denature probe DNA by placing tube in 90 °C water bath for 5 min.

^a If probe is labelled with digoxigenin-11-dUTP use method of detection as described in Chapter 9.

- 9** Plunge tube on ice.
- 10** Microfuge tube briefly to get all liquid to bottom of tube.
- 11** Place tube in 37 °C water bath for 2–3 h and prepare slides during this incubation.

(b) Slide preparation

Materials

- RNase A (100 µg ml⁻¹)

Method

- 12** Spread metaphases onto slides using standard procedures.
- 13** Place 100 µl RNase on slides and place coverslip on top.
- 14** Incubate slides in a humid chamber at 37 °C for 30–60 min.
- 15** Meanwhile, prepare slide-denaturing solution (see below) in Coplin jar and place in 75 °C water bath.
- 16** Remove coverslips after 1 h.
- 17** Wash slides twice in 2×SSC with agitation (3 min in each).
- 18** Dehydrate through ethanol series (3 min in each).
- 19** Air-dry slides.

(c) Denaturation of chromosomal DNA

Materials

- denaturing solution: 35 ml formamide, 5 ml 20×SSC (pH 5.3), 10 ml sterile distilled water

Method

- 20** Place slides in denaturing solution at 75 °C for 3 min.
- 21** Dehydrate through ice-cold ethanol series (3 min in each).
- 22** Air-dry slides.

(d) Hybridization

Additional material

- cowgum

Method

- 23** Prewarm slides in 42 °C water bath for 2 min.
- 24** Remove probe from 37 °C water bath and apply to slide.
- 25** Cover slide with a 32×22 mm coverslip and seal edges with cowgum.
- 26** Incubate slides at 37 °C overnight.

(e) Posthybridization washes**Materials**

- 3×50-ml washes: 25 ml formamide, 5 ml 20×SSC (pH 5.3), 20 ml sterile distilled water
- 3×50 ml washes: 2×SSC (pH 7)

Method

- 27** Prewarm the six wash solutions to 42 °C and prepare blocking solution (see below).
- 28** Place slides in 2×SSC to loosen rubber solution.
- 29** Remove rubber solution carefully with forceps and soak slides in 2×SSC for 5 min to loosen coverslips.
- 30** Gently remove coverslips.
- 31** Wash slides in the three formamide washes at 42 °C (5 min in each).
- 32** Wash slides in the three 2×SSC washes at 42 °C (5 min in each).

(f) Detection of biotin label**Materials**

- blocking solution: 1.8 g BSA, 60 ml 4×SSCT (4×SSC + 0.05% Triton X-100)
- layers 1 and 3: 1 µl avidin-FITC^b (5 µg ml⁻¹), 99 µl filtered blocking solution
- layer 2: 1 µl biotin antiavidin^c (5 µg ml⁻¹), 99 µl filtered blocking solution
- Citifluor/PI: 1 ml citifluor mountant, 8 µl propidium iodide (50 µg ml⁻¹)

Method

- 33** Wash slides at room temperature in 4×SSCT with agitation.
- 34** Place slides in blocking solution for 10–20 min.
- 35** Wash slides for 3 min with agitation in 4×SSCT.

^b e.g. Fluorescein Avidin DCS (cell sorting grade) (Vector Laboratories).

^c e.g. Biotinylated antiavidin D (Vector Laboratories).

- 36 Wipe backs of slides with tissue but do not allow to dry out.
- 37 Apply 100 µl of layer 1 to slide and place coverslip on top.
- 38 Incubate slides at 37 °C for 30 min.
- 39 Wash slides three times in 4×SSCT (3 min in each).
- 40 Place 100 µl of layer 2 on slide, coverslip and incubate at 37 °C for 20 min.
- 41 Repeat step 39.
- 42 Repeat steps 36 and 37.
- 43 Wash slides in 4×SSCT for 3 min.
- 44 Wash slides twice in PBS (5 min in each).
- 45 Dehydrate through ethanol series.
- 46 Air-dry slides.
- 47 Mount slides in 40 µl citifluor/PI or alternatively citifluor/DAPI and coverslip.
- 48 View slides on confocal microscope or fluorescence microscope (see Chapter 13 for digital microscopy).

.....

Troubleshooting

(See also the Troubleshooting section in Chapter 9.)
The most commonly encountered problems in FISH are the absence of any hybridization signal or a high background signal.

No signal

If no signal is apparent even after additional amplification steps:

- *recheck the concentration of the DNA used for labelling;*
- *recheck the quality of biotin labelling using a dot-blot assay [25];*
- *use more probe DNA in FISH experiment.*

Lots of background signal

This may be the result of inadequate suppression of repetitive sequences:

- *increase the amount of Cot-1 DNA in the prehybridization step;*
- *ensure that the posthybridization washes are at the correct temperature.*

For probes with sequences homologous to other regions, you will need to increase the stringency conditions during hybridization or posthybridization washes.

- Decrease the salt concentration in the hybridization buffer (e.g. use 2×SSC rather than 4×SSC).
- *Increase the formamide concentration in the hybridization buffer from 50% up to a maximum of 70%.*
- *Increase the temperature at which hybridization is performed (try 42 °C rather than 37 °C).*
- *Perform higher stringency posthybridization washes using lower salt concentrations down to 0.1 × SSC.*

Slides should never be allowed to dry out during the experiment.

Poor chromosome morphology

Fuzzy or swollen chromosomes may represent poor harvesting and slide-making conditions or suboptimal conditions during FISH. Cytoplasm surrounding chromosomes is a major problem in FISH. It can lead to high background signals and prevent the probe hybridizing efficiently. If various spreading methods have been tried (see Troubleshooting for Protocol 58) and cytoplasm is still a problem, then one or more of the following may be useful.

- *Postfix slides in formaldehyde or acetone for 10 min.*
- *Place slides in 0.2 M HCl for 20 min.*
- *Treat slides with proteinase K or detergent to permeabilize the cells.*
- *Place slides in 100% glacial acetic acid and if cytoplasm is still present place in 70% glacial acetic acid.*

If the chromosomes are fuzzy it is likely that they have been overdenatured. Do not denature the chromosomes at temperatures above 75 °C.

If the propidium iodide staining is very bright this may reflect insufficient chromosomal denaturation.

Protocol 63 Microcloning of DOP-PCR amplification products

Materials

- purification column (Promega Wizard PCR prep column)
- pBluescript II SK (Stratagene)
- calf intestinal alkaline phosphatase (CIP)
- *Xho*I restriction enzyme
- DNA ligase
- competent cells (e.g. DH5α; Stratagene)
- LB-amp-X-gal agar plates

Method

- 1** Purify^a about 1 μg PCR product.

^aPurify using GeneCleaning (Bio 101) or by purification column (Promega Wizard PCR prep column).

^b Positive clones can be picked off to master plates or alternatively into microtitre plates.
^c Glycerol stocks should be made as a permanent resource.

- 2 Digest DNA with *Xho*I (5 U mg⁻¹ DNA).
- 3 Purify^a digested DNA.
- 4 Ligate with 100 ng *Xho*I-digested and phosphatased plasmid vector.
- 5 Transform *E. coli* (see ref. 15).
- 6 Identify positive (white) clones.^{b,c}

.....

Troubleshooting

Low numbers of white colonies

It should be possible to tell from the controls included in the ligation/transformation reaction whether these steps need to be optimized. If the ligation reactions appears to be the fault, then repeat the ligation with a range of vector:insert ratios.

The primer has been designed with an *Xho*I recognition site with sufficient 5' nucleotides to get efficient digestion. However, it is common to get poor digestion of sequences which lie very close to the 5' end of the PCR fragment. To determine whether this has occurred, repeat the ligation with a control *Xho*I fragment (e.g. some λ -DNA digested with *Xho*I). If this fails then the *Xho*I vector should be prepared again.

Too many blue colonies

Again, this may be due to suboptimal ligation/transformation conditions, which could be improved by:

- *repreparing pBluescript with a new phosphatase;*
- *adding more PCR product insert.*

However, we have found that 'light blue' colonies frequently contain inserts, either small inserts or in-frame inserts which allow for some readthrough of the β -galactosidase.

.....

Protocol 64

Colony PCR to isolate microclone inserts

Materials

- reaction mixture: 5 μ l 10 \times *Taq* polymerase buffer, 2 μ l MgCl₂ (50 mM stock), 1 μ l dNTPs (5 mM stock), 0.5 μ l 5' vector primer (20 μ M stock), 0.5 μ l 3' vector primer (20 μ M stock), 0.5 μ l *Taq* polymerase (5 U μ l⁻¹), 40 μ l sterile H₂O, mineral oil, colony

Method

- 1 Place reaction mixture into sterile Eppendorf tube and inoculate with a colony.
- 2 Overlay with a drop of mineral oil and amplify using the following thermal cycle programme.
94 °C for 5 min;
1 cycle of:
94 °C for 45 s;
55 °C for 5 min;
72 °C for 5 min;
30 cycles of:
94 °C for 1 min;
55 °C for 1 min;
72 °C for 1 min;
and then:
72 °C for 10 min.

.....
Troubleshooting

We have found colony PCR to be relatively foolproof, and achieved single bands from about 99% of colonies tested from the microclone library we constructed. The following are some minor problems that we encountered.

No band

- Repeat PCR, one of the reaction components may have been accidentally left out.
 - Repeat PCR with different 5' and 3' vector primers.
 - Repeat PCR with increased amount of primer and Taq polymerase.
 - Hot-start PCR.
- If all these have failed, then it is possible that the clone contains a large insert that is too large to get efficient amplification under the conditions used. We have efficiently amplified fragments as large as 1600 base pairs. For larger inserts a miniprep would provide sufficient DNA for analysis.

More than one band

The colony is likely to be contaminated with a second colony. Restreak the colony, isolate several individual clones and re-check by colony PCR.

Smear of DNA

Repeat but inoculate with less colony. We have found that the smallest amount of colony is sufficient for efficient amplification.

.....

Acknowledgements

The authors thank Dr Peter Riddle who has been inspirational in setting up the microdissection technique, Mary Cotter for her secretarial assistance and Professor Bryan Young and Dr Trivadi Ganesan for their continued enthusiasm.

References

1 Scalenghe, F., Turco, E., Edstrom, J.E. *et al.* (1981) Microdissection and cloning of DNA from a specific region of *Drosophila melanogaster* polytene chromosomes. *Chromosoma* **82**, 205–216.

2 Guan, X.Y., Trent, J.M. & Meltzer, P.S. (1993) Generation of band-specific painting probes from a single microdissected chromosome. *Hum. Mol. Genet.* **2** (8), 1117–1121.

3 Hadano, S., Watanabe, M., Yokoi, H. *et al.* (1991) Laser microdissection and single unique primer PCR allow generation of regional chromosome DNA clones from a single human chromosome. *Genomics* **11**, 364–373.

4 Edstrom, J.-E., Kaiser, R. & Rohme, D. (1987) Microcloning of mammalian metaphase chromosomes. *Meth. Enzymol.* **151**, 503–516.

5 Meltzer, P.S., Guan, X.Y., Burgess, A. & Trent, J.M. (1992) Rapid generation of region specific probes by chromosome microdissection and their application. *Nature Genet.* **1**, 24–28.

6 Kaiser, R., Weber, J., Grzeschick, K.-H. *et al.* (1987) Microdissection and microcloning of the long arm of human chromosome 7. *Mol. Biol. Repts* **12**, 3–6.

7 Rooney, D.E. & Czepulkowski, B.H. (1992) *Human Cytogenetics: A Practical Approach*, 2nd edn (The Practical Approach Series, Vols 1 & 2, Rickwood, D. & Hames B.D., eds) (IRL Press, Oxford University Press).

8 Telenius, H., Carter, N.P., Bebb, C.E. *et al.* (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* **13**, 718–725.

9 Carter, N.P., Ferguson-Smith, M.A., Perryman, M.T. *et al.* (1992) Reverse chromosome painting: a method for the rapid analysis of aberrant chromosomes in clinical cytogenetics. *J. Med. Genet.* **29**, 299–307.

10 Rack, K.A., Harris, P.C., MacCarthy, A.B. *et al.* (1993) Characterisation of three de novo derivative chromosomes 16 by ‘reverse painting’ and molecular analysis. *Am. J. Hum. Genet.* **52**, 987–997.

11 Lüdecke, H.J., Senger, G., Claussen, U. & Horsthemke, B. (1989) Cloning defined regions of the human

genome by microdissection of banded chromosomes and enzymatic amplification. *Nature* **338**, 348–350.

12 Saunders, R.D.C. (1990) Short cuts for genomic walking: chromosome microdissection and the polymerase chain reaction. *BioEssays* **12** (5), 245–248.

13 Johnson, D.H. (1990) Molecular cloning of DNA from specific chromosomal regions by microdissection and sequence-independent amplification of DNA. *Genomics* **6**, 243–251.

14 Kao, F.T. (1993) Microdissection and microcloning of human chromosome regions in genome and genetic disease analysis. *BioEssays* **15** (2), 141–146.

15 Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn (Cold Spring Harbour Press, Cold Spring Harbour, New York).

16 Yu, J., Tong, S., Yang-Feng, T. & Kao, F.T. (1992) Construction and characterisation of a region-specific microdissection library from human chromosome 2q35-q37. *Genomics* **14**, 769–774.

17 Brown, S.D.M. & Greenfield, A.J. (1987) A model to describe the size distribution of mammalian genomic fragments recovered by microcloning. *Gene* **55**, 327–332.

18 Feinberg, A.P. & Vogelstein, B. (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Addendum. *Anal. Biochem.* **137**, 266–267.

19 Church, G.M. & Gilbert, W. (1984) Genomic sequencing. *Proc. Natl Acad. Sci. USA* **81**, 1991–1995.

20 Hirst, M.C., Bassett, J.H.D., Roche, A. & Davies, K.E. (1992) Preparation of radiolabelled hybridisation probes by STS labelling. *Trends Genet.* **8** (1), 6–7.

21 Cole, C.G., Dunham, I., Coffey, A.J. *et al.* (1992) A random strategy for construction of YAC contigs spanning defined chromosomal regions. *Genomics* **14**, 256–262.

22 Zhang, J., Meltzer, P., Jenkins, R. *et al.* (1993) Application of chromosome microdissection probes for elucidation of BCR-ABL fusion and variant Philadelphia chromosome translocations in chronic myelogenous leukemia. *Blood* **81**, 3365–3371.

23 Cotter, F.E., Lillington, D.M., Hampton, G. *et al.* (1991) Gene mapping by microdissection and enzymatic amplification: heterogeneity in leukaemia associated breakpoints on chromosome 11. *Genes Chrom. Cancer* **3**, 8–15.

24 Lillington, D.M., Cotter, F.E. & Riddle, P.N. (1992) Microdissection of human chromosomes. In *Human Cytogenetics: A Practical Approach* (Rooney, D.E. & Czepulkowski, B.H., eds), Vol. 2. *Malignancy and Acquired Abnormalities* (IRL Press, Oxford).

- 25 Lichter, P. & Cremer, T. (1992) Chromosome analysis by non-isotopic *in situ* hybridization. In *Human Cytogenetics: A Practical Approach* (Rooney, D.E. & Czepulkoski, B.H., eds), Vol. 1. *Constitutional Analysis*, 2nd edn (IRL Press, Oxford).
- 26 Yokoi, H., Hadano, S., Kogi, M. *et al.* (1994) Isolation of expressed sequences encoded by the human Xq terminal portion using microclone probes generated by laser microdissection. *Genomics* **20**, 404–411.

Chapter 12

Chromosome sorting and analysis by fluorescence-activated flow sorting

Simon P. Monard

Aaron Diamond AIDS Research Center, New York, NY10016, USA

12.1 Introduction, 292

12.2 Instrumentation, 296

Troubleshooting, 297

Poor discrimination between chromosomes during flow sorting, 297

12.3 Chromosome preparation, 298

12.4 Flow sorting chromosomes for library construction, 298

12.5 Generation of chromosome paints, 298

12.5.1 Degenerate oligonucleotide primed polymerase chain reaction, 298

12.5.2 Alu-polymerase chain reaction, 298

Protocol 65

Preparation of chromosomes by the polyamine method for flow sorting, 299

Protocol 66

Chromosome preparation by the magnesium sulphate method, 301

Troubleshooting, 301

Few chromosomes in the preparation and few intact cells at metaphase, 301

Many cells in metaphase but few chromosomes released into suspension, 302

Death of cells after colcemid treatment, 302

Protocol 67

Preparation of chromosomes for library construction, 302

References, 303

12.1 Introduction

Many types of cancer and genetic diseases are characterized by chromosomal aberrations. Conventional cytogenetics, the analysis of banded metaphase chromosomes (see Chapters 7 and 8), is a widely used technique in haematology, oncology and prenatal diagnosis to investigate such aberrations. But in cancer cytogenetics, where complex karyotypes are often encountered, such analysis can be very time consuming and in many cases marker chromosomes cannot be identified by their banding pattern alone.

There has therefore for some years been interest in a more rapid and objective method of analysing chromosomes. Flow cytometry offers an alternative machine-based approach to conventional cytogenetics. A suspension of metaphase chromosomes is prepared, stained with one or two DNA-binding fluorochromes and passed through a flow cytometer, in which the signal from each chromosome is measured and recorded as it passes through a focused laser beam (see refs 1 and 2 for a fuller description of the workings of a flow cytometer and flow cytometry in general). In the case of the single-laser flow cytometer, where chromosomes are stained with only one dye, such as ethidium bromide, which binds non-specifically to any DNA, the intensity of fluorescence from each chromosome is directly proportional to its DNA content. The data from 10 000–50 000 chromosomes is accumulated and presented as a histogram of fluorescence intensity against frequency. This plot shows a distinctive species-specific pattern of peaks and is called a univariate flow karyotype.

Not all the human chromosomes appear as separate peaks in a univariate plot, however. To achieve a complete resolution, chromosomes can be stained with two dyes that have a base-pair preference, and examined in a dual-laser flow cytometer. The dyes commonly used are Hoechst

33258 and chromomycin A₃. The Hoechst dye is an ultraviolet-excited fluorochrome which has an AT base-pair preference; chromomycin A₃ is excited by the blue light at the 457.9 nm line of an argon ion laser and has a CG binding preference. Chromosomes stained with these two dyes give a signal whose intensity is influenced not only by their DNA content but also by their base-pair composition. The intensity of each fluorescent signal is recorded for each chromosome; the data can be presented as two histograms. These can be combined to form an isometric plot (Fig. 12.1), or even better, presented as a dot-plot or contour map (Fig. 12.2). This bivariate flow karyotype resolves all human chromosomes as separate peaks except for chromosomes 9, 10, 11 and 12.

The flow karyotype provides no information about an individual cell but can provide an accurate measurement of the frequency of the different chromosome types. Trisomy 21, for example, would appear as a 50% increase in the frequency of chromosome 21 as compared with the other chromosomes. Translocations resulting in two derivative chromosomes that differ in either DNA content or base-pair ratio from the chromosomes from which they are derived will appear as two separate peaks in positions where there are normally none (Fig. 12.3). Small marker chromosomes and deletions can also usually be detected (Fig. 12.4).

Why then has this rapid, highly reproducible machine-based approach not completely displaced conventional cytogenetics? There are several reasons. Sometimes the derivative chromosomes appear in the same position as other chromosomes, making them difficult to detect. Also, a reciprocal translocation resulting in two derivative chromosomes which have the same DNA content and base-pair ratio as the parent chromosomes would remain undetected. Another reason is the expense of a dual-laser flow cytometer and the high level of expertise needed to operate it.

The major obstacle, however, is the polymorphic nature of the population, so that the two homologues of the same chromosome type in a normal individual often appear as separate peaks (Fig. 12.4). Therefore, in practice it would be difficult to ascertain whether two peaks seen in a flow karyotype were the two normal homologues of a chromosome or whether one was abnormal. Certain chromosomes also have regions of centric heterochromatin which can vary considerably in size, resulting in microscopically visible differences [3] and differences in the flow karyotype.

Variations in flow karyotypes have been correlated with specific C- or quinacrine-band poly-

Chromosome sorting and analysis by FACS are used for:

- the analysis of chromosome suspensions for the detection and measurement of chromosome aberrations
- sorting of the different chromosomes for the preparation of chromosome-specific DNA libraries
- the identification and characterization of marker chromosomes using reverse chromosome painting
- the preparation of chromosome-specific probes
- the isolation of single chromosomes

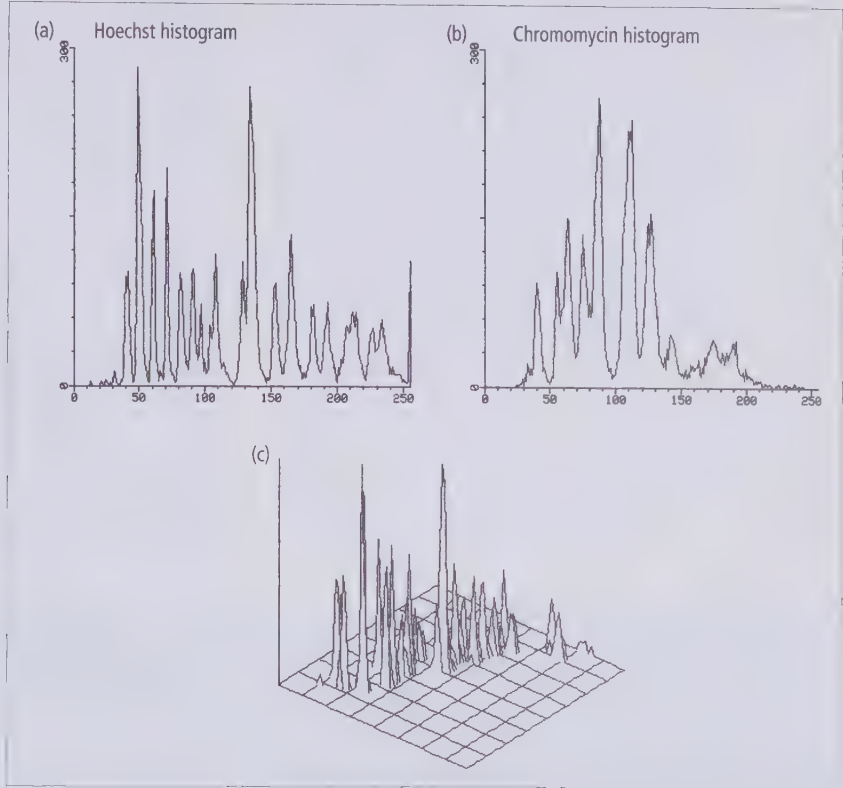


Fig. 12.1 Separate histograms and isometric plot of a normal female lymphoblastoid cell line. (a) Chromosomes stained with Hoechst 33258; (b) chromosomes stained with chromomycin A₃; (c) combined isometric plot of data from (a) and (b).

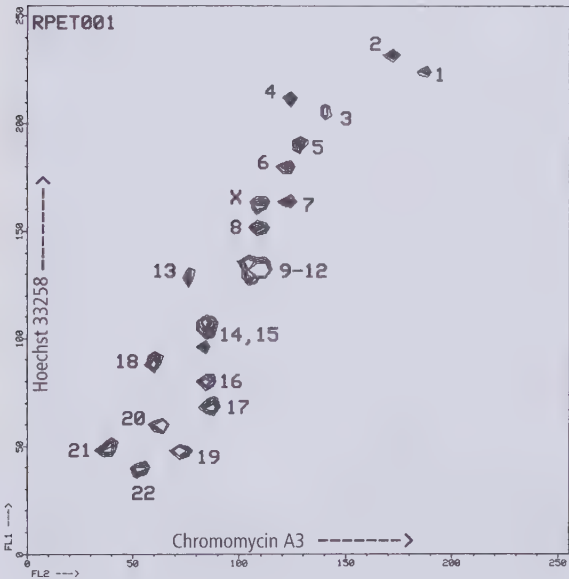


Fig. 12.2 Bivariate flow karyotype of the sample displayed in Fig. 12.1. One homologue of chromosome 15 appears in the same peak as chromosome 14.

morphisms [4, 5]. As chromosome polymorphisms appear to be inherited unaltered in size [6], any feature that cannot be seen in either of the parental flow karyotypes can be assumed to have arisen *de novo* [7]. This approach has been used to study a series of families with dysmorphic children [8]. Although in this study no chromosome abnormality

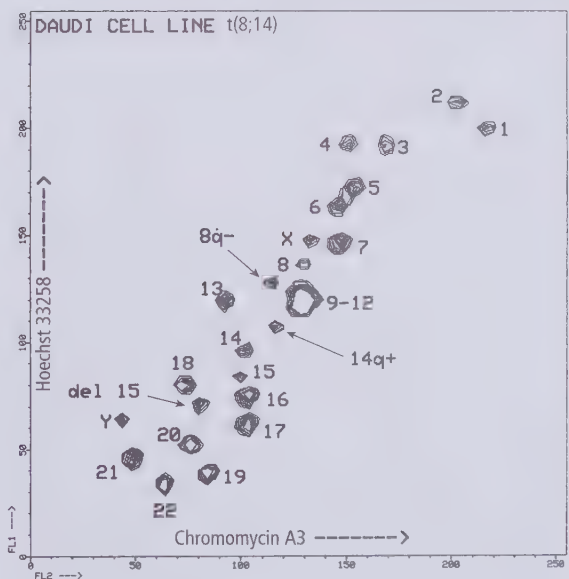


Fig. 12.3 Bivariate flow karyotype of Daudi cell line which carries the t(8;14) translocation found in Burkitt's lymphoma. The two translocation products (8q- and 14q+) are identified. One of the chromosome 15 homologues has a deletion and appears as a separate peak (del 15).

was detected, the usefulness of this approach was clearly demonstrated. Such family studies would prove difficult in cancer cytogenetics where the age of presentation is often in late middle age, and the

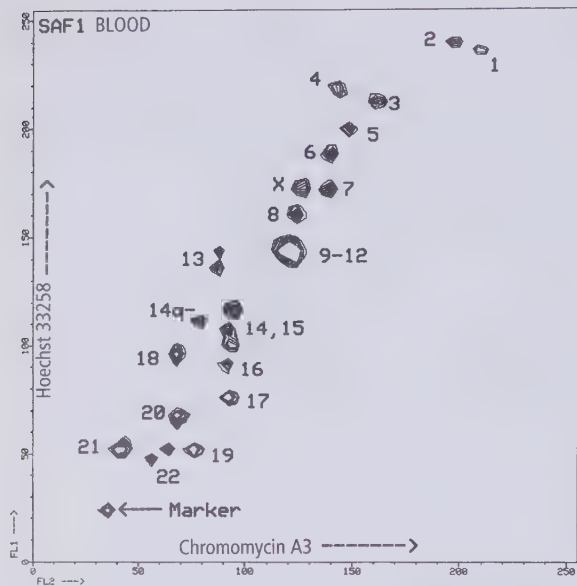


Fig. 12.4 Bivariate flow karyotype of phytohaemagglutinin (PHA)-stimulated peripheral blood culture from a female with a psychodevelopmental disorder. The picture is normal except for a deletion of one chromosome 14 (14q-) and a small marker which is about half the size of a G group chromosome. The two homologues of chromosome 22, although normal, appear as separate peaks. Reverse chromosome painting was used to demonstrate the marker to be composed largely of chromosome 14 material.

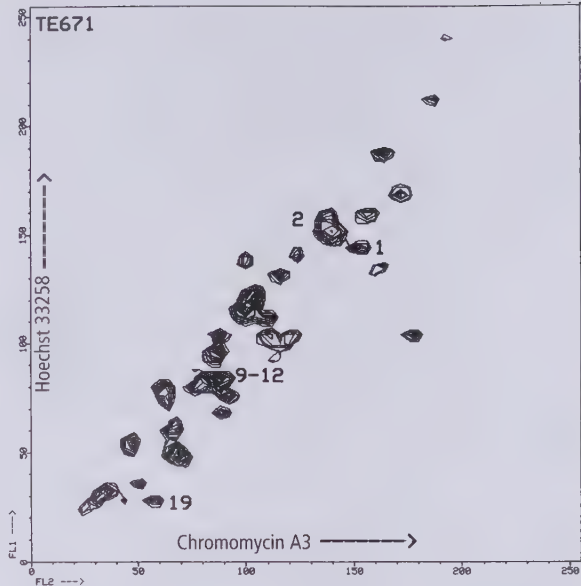


Fig. 12.5 Bivariate flow karyotype of a cell line derived from a patient with rhabdomyosarcoma. The chromosomes are highly rearranged, so much so that it is difficult to decipher which peaks are representing the normal chromosomes. Chromosomes 1, 2, 9–12 and 19 are indicated.

need to perform family studies is hardly a step towards a rapid machine-based approach to chromosome analysis. As with conventional cytogenetics, the very complex aberrations seen in some solid tumours such as rhabdomyosarcomas are difficult to interpret (Fig. 12.5).

The first human flow karyotypes were demonstrated in the mid-1970s using chromosomes isolated from fibroblasts [9, 10]. High-resolution flow karyotypes were subsequently obtained from phytohaemagglutinin (PHA)-stimulated peripheral blood lymphocytes [4] and lymphoblastoid cell lines [11] (see Fig. 12.2). In practice, chromosomes can be prepared from almost any culture of growing cells.

The real power of the flow system is the ability of the flow cytometer to separate the different chromosome types physically. Any chromosome that can be resolved as a separate peak can thus be sorted with a high degree of purity: a purity of 95% is typical and up to 99% purity is possible. Phage libraries [11] and, more recently, cosmid libraries [12], have been constructed from flow-sorted chromosomes. These libraries have proved to be an essential resource for the study of the human genome.

Suspension cell lines are the most convenient type of cells for preparing very large numbers of chromosomes, for example for sorting for chromosome library construction. Cells grown as mono-

layers, on the other hand, produce chromosome preparations with less debris, as dead cells can be removed by washing. It should be remembered that transformed cell lines, especially rodent cell lines, often develop chromosome aberrations that were not present in the primary tissue from which they were derived.

Sorting chromosomes for library construction is time consuming. It may take several days of sorting before enough chromosomes are separated. More appealing to the flow cytometrist is the combination of the polymerase chain reaction (PCR) [13] and flow sorting. Sorting enough chromosomes for a PCR reaction takes only a few seconds. If Alu-PCR [14] or degenerate oligonucleotide primed PCR (DOP-PCR) [15] is used to amplify chromosomal DNA, a library of PCR products will be produced ranging from about 500bp to 3kb in length and all having sequence identity to the chromosome from which they were generated (see Chapters 9–11 for protocols for Alu-PCR and DOP-PCR).

These PCR products can be used in several ways. They can be used as chromosome-specific probes for the purpose of isolating cosmids or yeast artificial chromosomes of interest [16], or they can be labelled with biotin-dUTP or fluorescent dUTPs and used as probes for fluorescence *in situ* hybridization [15,17] (see Chapters 9–11). The use of such fluorescent probes directed against the whole chromosome is termed chromosome painting (Chapter 10) and the probes are referred to as chromosome paints. The

resultant signal identifies the chromosome type from which the paint was prepared. Chromosome paints prepared from normal chromosomes and applied to metaphases containing abnormal chromosomes can reveal the identity of the aberrant chromosomes; this is termed forward chromosome painting.

Chromosome paints can be prepared from any of the chromosomes that resolve as separate peaks in flow cytometry. The positions of the different human chromosomes in the bivariate flow karyotype are well established, so paints can readily be prepared from all the human chromosomes except chromosomes 9–12.

A more rapid way of identifying the make-up of an unidentified marker chromosome is to flow sort that chromosome, prepare a chromosome paint and

apply this paint to a metaphase spread from a normal individual. The signal will be seen to be restricted to those chromosomes from which the marker is composed. This technique has been termed reverse chromosome painting. Chromosome paints have also been produced in other species. In the pig, the positions of the different chromosomes in the flow karyotype have been established by *in situ* hybridization and the paints have been used to investigate chromosome aberrations [18].

Chromosome paints can be used for the detection of structural and numerical aberrations in tissue preparations and cell suspensions fixed on slides [19,20]. One potentially important application of these techniques is the evaluation of the toxicological properties of chemical substances [21] (see Case Study 12.1).

Use of flow-sorted rat chromosomes to generate chromosome-specific paints for use in genotoxicity assays

In the toxicological evaluation of chemical substances, one of the internationally accepted tests used to determine the potential genotoxic properties of chemical compounds is the *in vivo* chromosomal aberration test (*OECD Guidelines for Testing of Chemicals*, Test 475). In this test chromosomal aberrations are scored in metaphases from bone marrow cells of treated animals. A significant drawback to this classical metaphase cytogenetic test is the requirement that cells have to be cultured *in vitro*, which is labour intensive, time consuming and often results in the selection of cells that grow well in culture but may not be representative of the original cell population. One solution to this problem is the use of the *in situ* hybridization (ISH) technique. ISH enables the detection of certain structural and numerical aberrations directly *in situ*—for example, in tissue preparations or cell suspensions fixed on slides [19,24,25].

The rat can be used as the test animal in the granuloma pouch assay (GPA). This *in vivo* assay can be used for the comparison of a variety of biological endpoints, such as HGPRT gene mutations, induction of DNA adducts, chromosome aberrations and induction of tumours (fibrosarcomas) in one and the same target tissue [26,27]. The goal of these studies was the development of a structural and numerical chromosome aberration test using the ISH technique in the GPA and other rat test systems.

A prerequisite for this approach is the availability of chromosome-specific probes. Although such probes are available for all the human chromosomes, very few are available for experimental animals such as the rat and the mouse. We therefore decided to isolate rat chromosome-specific probes ourselves using flow sorting and PCR.

Isolation of chromosomes from cell lines was ruled out

because of the spontaneous chromosome aberrations that occur in such cultures, so initially we isolated chromosomes from readily available rat splenocytes, stimulated with the mitogen concanavalin A (Con A). The bivariate flow karyotype looked promising in that many of the chromosomes resolved as separate peaks. We then isolated chromosomes from the primary fibroblasts generated in the GPA assay. These cells grow rapidly for several weeks and grow as a monolayer. The advantage of such cells over Con A-stimulated splenocytes is that dead cells and cell debris are removed when the cells are passaged, resulting in superior resolution, particularly of the smaller chromosomes.

Chromosomes were isolated using the polyamine method described in Protocol 65. Between 500 and 1000 of all the chromosome types that could be resolved were sorted directly into PCR tubes, the DNA was amplified using DOP-PCR and labelled with biotin either using nick translation or PCR amplification [15]. Chromosome-specific paints were generated against chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 19, 20, X and Y. Chromosomes 11, 13, 14 and 15 appeared as a single peak, as did chromosomes 16, 17 and 18, so individual paints could not be generated for these chromosomes. Subsequent chromosome painting experiments demonstrated the specificity of these probes.

Generating these rat chromosome probes was fairly rapid and straightforward. It required culturing cells for a few days, preparing the chromosome suspensions, which takes about an hour, staining the suspension and sorting the chromosomes into PCR tubes, which again took about an hour. The PCR was performed overnight. An alternative approach would be to generate radiation hybrid cell lines with single rat chromosomes in a mouse or hamster background, which would take many weeks.

An identical approach could be used to generate chromosome painting probes for other species.

Flow sorting has been used to generate probes against rat chromosomes for use as chromosome paints in the development of a rat model of carcinogenesis and mutagenesis. Probes were rapidly generated using a combination of flow sorting and PCR [21]. Chromosome-specific probes against 15 of the 22 different rat chromosomes were generated in about two days.

Chromosome paints generated by Alu-PCR do not paint the chromosome evenly; a series of bright and dark bands is seen which corresponds to the reverse-banding pattern. This is due to the dark bands produced by Giemsa staining having few Alu repeats. More even painting is achieved using probes generated by DOP-PCR. The primer used in this PCR reaction, unlike Alu primers, is not species dependent and produces a signal that is more evenly distributed over the chromosome, although, as with the Alu-PCR generated paints, the centromere is not usually painted.

The PCR products can be labelled in two main ways.

1 Labelling during the PCR reaction Biotin-11-dUTP can be incorporated during the PCR reaction and avidin-fluorescein isothiocyanate (FITC) used to develop the signal. The signal can be further amplified using FITC-labelled anti-FITC antibodies. Fluorochrome-labelled dNTPs can be incorporated into the PCR reaction producing directly labelled paints. Several suppliers produce dUTPs labelled with fluorescein, rhodamine and coumarin (see Appendix III). In general, the higher the proportion of labelled dNTP used, the lower the yield of the reaction in terms of micrograms DNA.

2 Labelling after the PCR reaction This can be done using one of the commercially available nick translation kits or random prime labelling kits (see Chapter 9, Protocol 44 and Chapter 10, Protocol 50).

12.2 Instrumentation

Univariate chromosome analysis can be performed on most flow sorters equipped with an argon ion laser. The computer system should allow the histogram to have at least 256 channels. Chromosomes are usually stained with ethidium bromide, which can be excited with the 514 nm line, or more usually with the 488 nm line, of the argon ion laser. The emission signal can be collected using a 580 nm long-pass filter.

Other fluorochromes can be used for univariate analysis and sorting if they give a clearer separation of the chromosome peak of interest. The choice of fluorochrome often depends on the light source available. Bench-top analysers with small air-cooled

lasers such as the FACScan (Becton Dickinson) cannot resolve the separate peaks sufficiently well to be useful for chromosome analysis. In our experience the best univariate flow karyotypes are obtained with Hoechst 33258.

Bivariate chromosome analysis and sorting using Hoechst 33258 and chromomycin A₃ requires a flow cytometer equipped with two argon ion lasers. The primary laser, that is the laser that intersects the sample stream nearest the nozzle, is tuned to the UV lines from 351.1–363.8 nm and is used to excite the Hoechst dye. The secondary laser is tuned to 457.9 nm and is used to excite the chromomycin. The signal from the Hoechst dye is collected using a 390 nm long-pass and a 480 nm short-pass filter. The chromomycin signal is collected using a 490 nm-long pass filter only. All filters are of the coloured glass variety. The signal from the primary laser is usually designated fluorescence 1 (FL1) and that from the secondary laser fluorescence 2 (FL2).

In our laboratory two dual-laser Becton Dickinson flow cytometers, a FACStar^{PLUS} and a FACS 440, are both used for chromosome sorting and analysis. With these instruments the two signals from the different dyes are separated in three ways.

1 Temporal separation A particle is illuminated by the secondary laser about 20 μ s after it is illuminated by the primary laser, so the signal from the chromomycin is collected 20 μ s after the instrument has been triggered by the Hoechst signal.

2 Spatial separation The primary laser strikes the stream above the secondary laser, the fluorescent signals are inverted by the collection optics and the secondary signal is reflected by a half mirror into the FL2 channel.

3 Optical separation by coloured glass filters in front of the photomultiplier tubes (PMTs) The filters in front of the FL1 PMT allow only blue light from 390 nm to 480 nm through. The FL2 signal is collected above 490 nm.

These three methods of separation ensure that the signals from the two dyes are measured completely independently from one another.

The alignment of the laser beams is of critical importance. Special care should be taken to ensure that they do not pass too close to the edge of any of the prisms as this can cause diffraction and subsequent loss of resolution. It is important that the lasers are functioning in the TEM₀₀ mode. The laser focusing lens is usually an achromatic doublet, that is two lenses of different materials positioned close to each other or cemented together. If the lens is of the cemented type, the cement may discolour after prolonged use in the UV and so should be inspected periodically.

The usual nozzle orifice size used is 50 μm although a 70- μm nozzle can be used. A dirty nozzle can cause poor resolution, increased noise and irregular deflection streams. When sorting chromosomes, particularly for library construction, keeping the sample and sorted fraction cool with ice or circulating cold water reduces the chance of DNA degradation.

Bivariate chromosome analysis can be performed by measuring the fluorescence parameters FL1 and FL2 alone, but it can be useful to use one or more scatter parameters for gating-out debris. The instrument should be triggered on the Hoechst signal and the minimum possible threshold should be used, as a high threshold would allow small particles such as chromosome fragments to pass undetected through the instrument and into the sorted droplets.

It is advisable to do a test sort before commencing any chromosome sorting. The test can be done on fluorescent microspheres or, preferably, on chromosomes. The sorted chromosomes can either be restained with Hoechst and chromomycin A₃ and rerun on the same instrument, or, if a bench-top analyser such as a FACScan is available, the

chromosomes can be stained with propidium iodide and analysed on this instrument. Bench-top cytometers, although unable to identify sorted chromosomes, are able to indicate whether the sorted fraction is composed of a single chromosome type and hence whether the instrument is sorting efficiently. If the sorted sample is restained and run on the same instrument, care should be taken that the sample tubing is thoroughly flushed through with sheath buffer to remove chromosomes adhering to the inside. It may be necessary to replace the sample tubing.

Sorting chromosomes for library construction is a time consuming process. With a good preparation and a well set-up instrument it should be possible to pass 1500–2500 chromosomes per second though the instrument without severe deterioration of resolution. Thus it should be possible to sort a single copy chromosome at a rate of 30–50 per second. When running chromosomes for analysis the best discrimination between chromosome peaks will be obtained using a low sample rate.

Use a sheath buffer containing 100 mM NaCl, 10 mM Tris HCl, 1 mM EDTA (pH 8) in distilled water.

.....

Troubleshooting

Poor discrimination between chromosomes during flow sorting

It is not always easy to determine if poor discrimination between chromosomes by a cytometer is due to a poor chromosome preparation, or poor cytometer or laser performance. The instrument can be monitored using fluorescent microspheres. The coefficients of variation of the fluorescent peaks and the intensity of the signal give an indication of how the instrument is performing.

Staining the chromosome preparation with propidium iodide and viewing with a fluorescence microscope will show whether there are too few chromosomes or the chromosomes are aggregating.

Finally, it is generally most useful to exchange preparations with another laboratory experienced in chromosome sorting and analysis.

Caution: Care should be exercised when aligning the lasers, especially in the UV. The lowest laser powers should be used and protective goggles must be worn.

.....

12.3 Chromosome preparation

Chromosome suspensions are prepared by adding an agent such as colcemid or vinblastine to a culture of growing cells, to arrest mitoses in metaphase and incubating at 37 °C for several hours or overnight until sufficient cells have accumulated in mitosis. Leaving for too long will result in death and necrosis of some cells which will produce DNA debris indistinguishable from chromosomes. Incubating for too short a period results in too few cells in mitosis. The ideal length of time depends on the rate at which the cells are growing. The key to making a good chromosome preparation is to start with a cell culture of healthy cells which are growing optimally. If the cells grow as an attached layer, mitotic shake-off can be used to obtain an enriched population of metaphase cells.

The electronics of commercially available flow sorters can only deal with flow rates of up to about 5000 events per second; if only a small proportion of these events are chromosomes then the actual number of chromosomes sorted will be low. For instance, ignoring abort and coincidence rates, at a sample rate of 5000 events per second, assuming all events are chromosomes, a single-copy chromosome will be isolated at a rate of about 100 per second. If only 10% of the events are chromosomes, for the same sample rate only 10 chromosomes a second will be isolated.

Sufficient chromosomes for flow karyotyping can be obtained from as little as 2ml of human peripheral blood. The use of short-term cultures has the advantage that there is little opportunity for karyotype alterations, which can occur in established cell lines. Also, the flow karyotype can be compared directly with the conventionally banded karyotype normally performed for routine analysis. Lymphoblastoid cell lines, on the other hand, represent an ideal source if large numbers of chromosomes are required, such as for sorting for library construction. Some lymphoblastoid cell lines carry known aberrant chromosomes that may be sorted to facilitate analysis of the genotype of the aberration or to simplify subchromosome gene mapping with known karyotype abnormalities.

There are two main methods used for the preparation of chromosomes for flow sorting: one uses polyamines to stabilize chromosomal DNA, the other uses magnesium sulphate. The polyamine method [22] is in routine use in our laboratory. The method was first described in 1981 and represented a breakthrough, as it enabled chromosome sorting and analysis to be performed on commercially available instruments with 5 watt lasers. The

preparations can be used for several weeks or even months, the DNA is of high quality and the resolution is good. We use the same protocol for preparing chromosomes from human and rodent cells.

It offers good discrimination between the chromosome types and the DNA after sorting is of very high molecular weight. This method is described in detail in Protocol 1. The magnesium sulphate method [23] offers excellent discrimination between the chromosomes and is rapid and simple to perform. The DNA, however, may not be of such good quality; this method is described in Protocol 66.

12.4 Flow sorting chromosomes for library construction

Cosmid libraries constructed from flow-sorted chromosomes [12] have proved a vital resource for the analysis of the human genome. Flow sorting for such a purpose can be time consuming so it is essential to sort at the highest rate possible without compromising purity. Keeping both the sample and sorted fractions cool reduces the chance of DNA degradation. A concentrated chromosome suspension with few interphase nuclei allows a higher sample rate for the same sample pressure. Preparation of chromosomes for library construction is described in Protocol 67.

12.5 Generation of chromosome paints

12.5.1 Degenerate oligonucleotide primed polymerase chain reaction

Degenerate oligonucleotide primed polymerase chain reaction (DOP-PCR) using the primer 6-MW can be used to generate chromosome paints from flow-sorted chromosomes from any species [15]. The six specific bases at the 3' end of the oligonucleotide prime theoretically every 4kb along the template DNA at the low annealing temperature. Only the oligonucleotide-'tailed' DNA generated in the initial cycles is amplified in the later high annealing temperature cycles. The paints generated using this primer 'paint' the chromosome evenly along its length although the centromere does not usually label.

A method for DOP-PCR is described in Protocol 61.

12.5.2 Alu-polymerase chain reaction

It may sometimes be desirable to generate paints

using primers directed against repeat sequences such as Alu repeats. Alu primers can be used to generate paints from human chromosomes in somatic cell hybrids where only the human material is amplified. Alu repeats occur on average about every 4kb in the genome but they are not evenly spaced. Painting using Alu paints results in uneven painting along the length of the chromosome giving a pattern corresponding to R-banding. There is a risk

that the region of interest will not be amplified when Alu paints are used; on the other hand the DOP-PCR paints do not paint entirely evenly, in the same way as with Alu paints, the centromere is rarely painted. Alu paints give low background and the pattern along the length of the chromosome can aid chromosome identification. Methods for Alu-PCR are given in Protocol 43, and Protocol 51.

.....

Protocol 65

Preparation of chromosomes by the polyamine method for flow sorting

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- colcemid solution (supplied as 10 µg ml⁻¹ solution) (Life Technologies)
- hypotonic solution (75 mM KCl)
- chromosome isolation buffer 1 (CIB1): 20 mM NaCl, 80 mM KCl, 15 mM Tris-HCl, 0.5 mM EGTA, 2 mM EDTA, 0.15% w/v 2-mercaptoethanol, 0.2 mM spermine (free base), 0.5 mM spermidine (free base), pH 7.2, in autoclaved distilled water
- digitonin (supplied as powder) (Sigma)
- propidium iodide (PI) 50 µg ml⁻¹ in phosphate-buffered saline (PBS)
- Hoechst 33258
- chromomycin A₃
- sodium citrate (100 mM)
- sodium sulphite (250 mM)
- bench centrifuge
- fluorescence microscope
- hotplate/magnetic stirrer
- incubator
- 12 × 75 mm plastic tubes

Method

- 1 Cell lines, either monolayer or suspension, may be used, as may PHA-stimulated peripheral blood cells. Whichever type of cell is used, the best preparations will be made from healthy cells growing optimally. Subculture cells 24 h before blocking with colcemid.
- 2 Block cells with 0.05 µg ml⁻¹ colcemid for 5–16 h depending on the rate of growth. Usually blocking overnight gives good results. The proportion of suspension cells in mitosis can be estimated by pelleting the cells from 1 ml of the blocked cell culture, discarding the

- supernatant and resuspending in PBS containing 50 and 0.1% Triton X-100. The estimation can be made either using a fluorescence microscope or a bench-top flow cytometer. It should be possible to get 40–60% of the cells in mitosis with suspension cell lines.
- 3** The proportion of monolayer cells in mitosis can be estimated on an inverted microscope. Mitotic cells are round and can usually be shaken off into the medium by giving the flask a sharp rap. Some monolayer cell lines may require the use of trypsin. Once in suspension centrifuge all types of cells at 100 *g* for 10 min in 50-ml plastic tubes, discard the supernatant and resuspend the cells in fresh medium before a further 10 min centrifugation at 100 *g*.
 - 4** Discard the supernatant by inverting the tube. Remove the last few drops from inside the tube with a tissue. Disaggregate the cell pellet by vortexing gently or by flicking the tube. Add 5 ml hypotonic solution, mix gently and leave for 10–30 min at room temperature (lymphoblastoid cell lines usually require 20 min, fibroblastoid cell lines usually require 30 min) . This is a convenient time to pool the contents of several tubes. Centrifuge the tubes for 10 min at 100 *g*.
 - 5** While the cells are in the swelling solution dissolve 12 mg digitonin in 5 ml distilled water by heating on a hotplate or in a microwave oven. Allow the digitonin solution to cool then add 1 ml 10×CIB1 and make the volume up to 10 ml with distilled water. Adjust the pH to 7.2 if necessary and place on ice.
 - 6** Following centrifugation, carefully remove the supernatant with a Pasteur pipette and agitate the tube gently to disaggregate the cells. Add 10 times the volume of the cell pellet in cold CIB1, and aspirate gently with a Pasteur pipette. Mix a small amount of the preparation with an equal volume of PI (50 µg ml⁻¹ in PBS) and view with a fluorescence microscope. If the chromosomes are not monodispersed, aspirate the preparation more vigorously or vortex gently. Avoid vortexing too vigorously as it can result in chromosome damage.
 - 7** The chromosome suspension may be stored at 4 °C for several weeks with little deterioration of flow karyotype.
 - 8** Transfer 1 ml of the chromosome suspension into a 12×75 mm plastic test tube, add 30 µl Hoechst 33258 (100 µg ml⁻¹ in distilled water), mix immediately. Add 40 µl 15 mM MgCl₂ and 50 µl chromomycin A₃ (2 mg ml⁻¹ in ethanol), mix and leave the sample at 4 °C for 2 h in the dark.
 - 9** The chromosome profile can be improved if 100 µl sodium citrate (100 mM) and 100 µl sodium sulphite (250 mM) are added at least 15 min prior to running on the cytometer. Aggregates and intact nuclei in the sample can be removed by centrifuging at 200 *g* for 1 min, then transferring the supernatant to a new tube. Centrifugation selectively depletes the larger chromosomes so should be avoided if the flow karyotype is to be analysed.
-

Protocol 66

Chromosome preparation by the magnesium sulphate method

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- chromosome isolation buffer 2 (CIB2): 40 mM KCl, 5 mM Hepes, 10 mM MgSO₄, 3 mM DTT (pH 8), in autoclaved distilled water
- Triton X-100
- growth media
- colcemid (Life Technologies)
- bench centrifuge
- fluorescence microscope
- incubator
- vortex mixer
- tissue culture flasks
- 50-ml conical tubes
- 12×75 mm plastic tubes
- phase contrast microscope

Method

- 1 Prepare colcemid-blocked cells as with the polyamine method (see Protocol 65).
- 2 Centrifuge cells at 300 g for 10 min at room temperature, decant supernatant, draining tubes on an absorbent paper towel.
- 3 Add 1 ml CIB2–6×10⁵ cells, resuspend gently and incubate at room temperature for 10 min.
- 4 Add 0.1 ml Triton X-100 solution (2.5% in distilled water) and incubate on ice for 10 min. Vortex for 10–20 s to disrupt the cells and incubate at room temperature for 10 min (monitor using phase contrast microscopy).
- 6 Stain for bivariate analysis as in step 8 of Protocol 65.

Troubleshooting

Two main problems can occur when preparing chromosomes.

Few chromosomes in the preparation and few intact cells at metaphase

This is almost certainly a problem with cell culture, resulting in a poor growth of cells. There can be many reasons for this ranging from media problems to mycoplasma infection.

Many cells in metaphase but few chromosomes released into suspension

Some cell types are very resistant to lysis.

- *To overcome this problem, one can increase the swelling time in KCl, use more detergent and vortex more vigorously.*

Death of cells after colcemid treatment

Some types of cells seem to die when left in the presence of colcemid for long periods, resulting in a mucous-like pellet after swelling. One solution to this problem is to block the cells for a shorter time, for instance 2 h. With very slow-growing cells, one strategy is to synchronize the cells, monitor the stage of the cell cycle and block the cells as they approach metaphase. It is also possible to remove dead cells by spinning the cells through a density gradient after blocking, removing the cells at the interface then washing in PBS before swelling them in KCl.

.....

Protocol 67 Preparation of chromosomes for library construction

Materials

- sheath buffer: 100 mM NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 8, in distilled water
- tRNA (Life Technologies)
- proteinase K (BDH)
- stock solution 500 mM EDTA, pH 8
- stock solution 20% (w/v) *n*-lauroylsarcosine (sodium salt)
- sterile 1.5-ml conical tubes with screw caps

Method

- 1** Prepare sterile sheath buffer containing 500 µg ml⁻¹ tRNA. Dispense 50 µl of this solution into sterile 1.5 ml conical tubes and vortex vigorously to coat the inside of the tubes. They can be stored by freezing quickly on dry ice and stored at -20 °C.
- 2** Sort 5 × 10⁵ chromosomes into each tube.
- 3** Prepare a working solution of 250 mM EDTA with 10% (w/v) *n*-lauroylsarcosine. This solution is added to the sorted chromosome suspension to make a final concentration of 25 mM EDTA and 1% *n*-lauroylsarcosine. When using a 50 µm nozzle and a 3-drop deflection, 5 × 10⁵ chromosomes should occupy a volume of about 700–800 µl. Thus 70–80 µl working solution should be added.
- 4** Add 180 µg Proteinase K to each tube, vortex and incubate at 42 °C overnight.

5 The resulting DNA preparation can be stored at 4°C for many months.

References

- 1 Shapiro, H.M. (1988) *Practical Flow Cytometry* (2nd edn, Alan Liss, New York).
- 2 Watson, J.V. (1991) *Introduction to Flow Cytometry* (Cambridge University Press, Cambridge).
- 3 Trask, B., van den Engh, G., Mayall, B. & Gray, J.W. (1989) Chromosome heteromorphism quantified by high-resolution bivariate flow karyotyping. *Am. J. Hum. Genet.* **45**, 739–752.
- 4 Young, B.D., Ferguson-Smith, M.A., Sillar, R. & Boyd, E. (1981) High resolution analysis of human peripheral lymphocyte chromosomes by flow cytometry. *Proc. Natl Acad. Sci. USA* **78**, 7727–7731.
- 5 Langlois, R.G., Yu, L.C., Gray, L.W. & Carrano, A.V. (1982) Quantitative karyotyping of human chromosomes by dual beam flow cytometry. *Proc. Natl Acad. Sci. USA* **79**, 7876–7880.
- 6 Robinson, J.A., Buckton, K.E., Spowart, G. *et al.* (1976) The segregation of human chromosome polymorphisms. *Am. J. Hum. Genet.* **40**, 113–121.
- 7 Trask, B., van den Engh, G. & Gray, J.W. (1989) Inheritance of chromosome heteromorphisms analyzed by high-resolution bivariate flow karyotyping. *Am. J. Hum. Genet.* **45**, 753–760.
- 8 Harris, P., Cooke, A., Boyd, E. *et al.* (1987) The potential of family flow karyotyping for the detection of chromosome abnormalities. *Hum. Genet.* **76**, 129–133.
- 9 Gray, J.W., Carrero, A.V., Moore, I.D.H. *et al.* (1975) High speed quantitative karyotyping by flow microfluorometry. *Clin. Chem.* **21**, 1258–1262.
- 10 Gray, J.W., Carrano, A.V., Steinmetz, L.L. *et al.* (1975) Chromosome measurement and sorting by flow systems. *Proc. Natl Acad. Sci. USA* **72**, 1231–1234.
- 11 Krumlauf, R., Jeanpierre, M. & Young, B.D. (1982) Construction and characterization of genomic libraries from specific human chromosomes. *Proc. Natl Acad. Sci. USA* **79**, 2971–2975.
- 12 Nizetic, D., Zehetner, G., Monaco, A.P. *et al.* (1991) Construction, arraying and high density display of large insert libraries of the human chromosomes X and 21: their potential use as reference libraries of these chromosomes. *Proc. Natl Acad. Sci. USA* **88**, 3233–3238.
- 13 Saiki, R.K., Gelfand, D.H., Stoffel, S. *et al.* (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491.
- 14 Nelson, D.L., Ledbetter, S.A., Corbo, L. *et al.* (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl Acad. Sci. USA* **86**, 6686–6690.
- 15 Telinius, H., Pelmeur, A., Tunnacliffe, A. *et al.* (1992) Cytogenetic analysis by chromosome painting using DOP-PCR amplifies flow sorted chromosomes. *Genes Chrom. Cancer* **4**, 257–263.
- 16 Cotter, F.E., Das, S., Douek, E. *et al.* (1991) The generation of DNA probes to chromosome 11q23 by Alu-PCR on small numbers of flow sorted 22q-derivative chromosomes. *Genomics* **9**, 473–480.
- 17 Suijkerbuijk, R., Matthopoulos, D., Kearney, L. *et al.* (1992) Fluorescent in situ identification of human marker chromosomes using flow sorting and Alu element-mediated PCR. *Genomics* **13**, 355–362.
- 18 Langford, C.F., Telinius, H., Carter, N.P. *et al.* (1992) Chromosome painting using chromosome specific probes from flow sorted pig chromosomes. *Cytogenet. Cell Genet.* **61**, 221–223.
- 19 Hopman, A.H.N., Ramaekers, F.C.S. & Vooijs, G.P. (1990) Interphase cytogenetics of solid tumours. In *In situ Hybridization, Principles and Practice* (Polak, J.M. & McGee, J. O'D., eds) (Oxford University Press, Oxford).
- 20 Cremer, T., Lichter, P., Borden, J. *et al.* (1988) The detection of chromosome aberrations in metaphase and interphase tumour cells by in situ hybridization using chromosome specific library probes. *Hum. Genet.* **80**, 235–246.
- 21 Hoebee, B., de Stoppelaar, J.M., Suijkerbuijk, R.F. & Monard, S. (1993) Isolation of rat chromosome-specific paint probes by bivariate flow sorting followed by degenerate oligonucleotide primed-PCR. *Cytogenet. Cell Genet.* **66**, 277–282.
- 22 Sillar, R. & Young, B. (1981) A new method for the preparation of metaphase chromosomes for flow analysis. *J. Histochem. Cytochem.* **29**, 74–78.
- 23 van den Engh, G.J., Trask, B., Gray, J.W. *et al.* (1985) Preparation and bivariate analysis of suspensions of human chromosomes. *Cytometry* **6**, 92–100.
- 24 Cremer, T., Popp, S., Emmerich, P. *et al.* (1990) Rapid metaphase and interphase detection of radiation-induced chromosome aberrations in human lymphocytes by chromosomal suppression in situ hybridization. *Cytometry* **11**, 110–118.
- 25 Evans, J.W., Chang, J.A., Giaccia, A.J. *et al.* (1991) The use of fluorescence in situ hybridization combined with preparative chromosome condensation for the identification of chromosome damage. *Br. J. Cancer* **63**, 517–521.
- 26 Maier, P. (1984) The granuloma pouch assay. In *Chemical Mutagens, Principles and Methods for Their Detection* (De Serres, F.J., ed.), Vol. 9, 233–260 (Plenum, New York).
- 27 Mohn, G.R., Maier, P., van Kreijl, C.F. *et al.* (1989) Further development of the granuloma pouch assay as an in vivo/in vitro model for comparing genetic, oncogenic, pre-malignant and neoplastic changes in a mesenchymal tissue of the rat. In *Cell Transformation and Radiation-induced Cancer* (Chadwick, K.H., Seymour, C & Barnhart, B. eds), 155–162 (Institute of Physics Publishing, Bristol).

Digital microscopy for fluorescence in situ hybridization

Nigel P. Carter

The Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1RQ, UK

13.1	Introduction, 306	13.4.1	Ratio labelling, 313
13.2	Epifluorescence microscopy, 306	13.5	Fluorescent chromosome band enhancement, 313
13.2.1	Electronic cameras, 307	13.6	Obtaining hardcopy output, 314
13.2.2	Video cameras, 308		
13.2.3	CCD cameras, 308		
13.2.4	Colour CCD cameras, 310		
13.2.5	Imaging multiple fluorochromes with digital cameras, 310	13.7	Image data storage, 314
13.3	Laser scanning microscopy, 311		Acknowledgements, 314
13.4	Multiple probe detection, 312		References, 315

13.1 Introduction

Fluorescence *in situ* hybridization (FISH) is making an increasingly important contribution to genome analysis. DNA sequences of less than 1 kilobase (kb) long can be mapped onto metaphase bands [1], and sequences can be ordered at resolutions down to 50 kb in interphase nuclei [2–4] and less than 5 kb on extended chromatin preparations [5–7] (Chapter 9), providing data of direct relevance to the physical mapping of the genome of humans and other species. Similarly, chromosome-specific sequences and more complex probes such as whole-chromosome paints [8] (Chapters 10 and 12) are used widely for the identification of chromosomal abnormalities and for prenatal and preimplantation diagnosis [9,10]. The advantage of FISH over other non-isotopic *in situ* hybridization detection methods [11] lies in the use of multicolor techniques which allow the identification of multiple probes on appropriately counterstained target DNA [12–15] (Chapters 9 and 10).

The popularity of FISH is due in no small part to recent advances in *digital microscopy* and *image analysis*. Modern microscope objectives coupled to digital imaging systems provide instruments of a sensitivity and resolution suitable for detection of small, weakly fluorescent signals. Multiple fluorochrome images can be acquired quickly and stored permanently using the personal computer, and the digital image processed in many different ways to enhance the information gained. Two basic microscope systems are in current usage for FISH analysis: (i) an *epifluorescence microscope* with an electronic camera attached, or (ii) *laser scanning technology*.

This chapter reviews the principles of digital microscopy for FISH and the equipment available. The preparation of probes for FISH and microFISH is covered in Chapters 9–11, and *in situ* hybridization procedures for fluorescence microscopy in Chapters 9 and 10.

13.2 Epifluorescence microscopy

The elements of an epifluorescent microscope are

Digital microscopy for FISH is used to:

- acquire digital images of fluorescent hybridizations
- enhance the sensitivity of fluorescence image detection
- allow image enhancement and quantitative analysis
- allow the use of ratio-labelled probes
- allow easy and rapid archiving and retrieval of images

Applications box 13.1

shown in Fig. 13.1. In the epifluorescent microscope, the excitation light source is usually a high intensity mercury arc lamp, although xenon lamps and lasers have also been implemented for this purpose. An arc lamp consists of a glass envelope containing a gas or vapour at high pressure. An initial high voltage spark between two electrodes within the envelope forms a luminous plasma arc which is maintained by the application of a high current at low voltage. Arc lamps act as point light sources of nonuniform radiance (the highest intensity is at the arc cathode) and are intrinsically unstable and prone to wander and flicker. As such, they are not suitable for use in critical illumination arrangements. Most usually, Köhler illumination is used, where the spot of highest intensity formed at the cathode is placed at the focal point of the lamphouse condenser lens. The condenser lens is then utilized as an illuminated disc which is imaged onto the specimen to generate uniform illumination across the field of view. Such even illumination, as provided by properly adjusted Köhler optics, is particularly important for advanced FISH techniques which rely on quantitative aspects of the image.

While the mercury arc lamp emits light over a wide range of wavelengths, peaks of higher intensity occur which relate to the characteristic spectral lines of mercury (Fig. 13.2). The sensitivity of the microscope (i.e. the brightness of the fluorescence) is influenced by the intensity of the illumination. Increasing the luminance of the arc lamp will increase the intensity of fluorescence until photosaturation and bleaching at high light intensities become significant. However, the luminance of an arc lamp is not only affected by its wattage but also by the current density and electrode geometry.

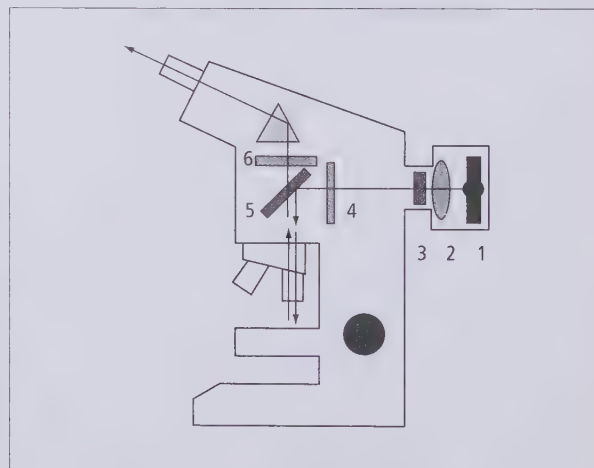


Fig. 13.1 The epifluorescence microscope. 1, mercury arc lamp; 2, condenser lens; 3, heat filter; 4, excitation filter; 5, dichroic mirror; 6, emission filter.

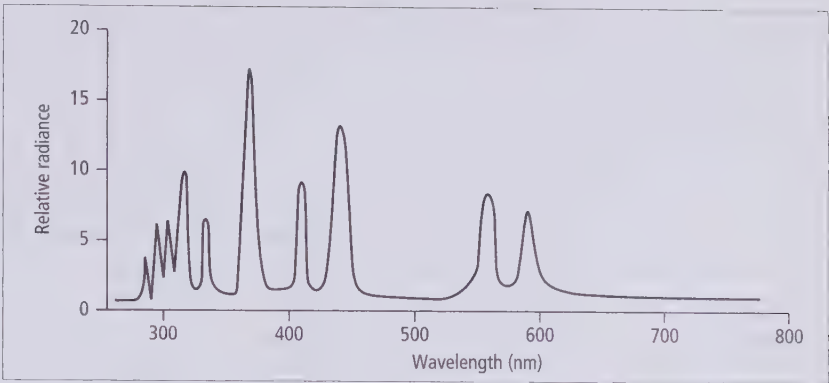


Fig. 13.2 Spectral radiance of mercury arc lamp. Redrawn from [18].

In practice, the 100 W compact mercury arc lamp provides the greatest source luminance (greater than the 200 W mercury arc lamp) and is the best choice of excitation source for FISH where the highest sensitivity is required. However, modern digital cameras are now very sensitive devices and do not require the highest illumination intensity to detect even weak fluorescent signals. For use with the most sensitive cameras, the 50 W compact mercury arc lamp is recommended due to its greater stability and uniformity of light distribution when used with Köhler illumination optics.

The objective lens in an epifluorescence arrangement is used both for illuminating the specimen with the excitation light as well as collecting the stimulated fluorescence. This dual purpose is made possible by an optical filter block consisting of an excitation filter, a dichroic mirror and an emission (barrier) filter. The excitation filter is used to select an appropriate wavelength for the excitation beam from the range of wavelength peaks emitted by the arc lamp. The dichroic filter reflects the excitation wavelength down on to the specimen while allowing longer wavelength fluorescence emission to pass through to the eyepieces. As the intensity of the excitation beam is many orders of magnitude greater than the stimulated fluorescence and optical filters are not 100% efficient, an emission filter is included in the imaging light path to prevent back-scattered excitation light reaching the eyepieces.

The numerical aperture of the objective lens is also an important parameter affecting the sensitivity of FISH. Numerical aperture is defined as $n \sin \theta$ where n is the refractive index of the medium between the specimen and the lens and θ is the half angle of the cone of light collected by the lens. The numerical aperture is thus a measure of the light-collecting properties of the lens. A high numerical aperture is achieved in objective lenses by using a combination of large aperture optical elements, reducing the distance from the front element of the lens to the specimen (the working distance) and by increasing

the refractive index between the specimen and the lens (e.g. oil immersion). In theory, the larger the numerical aperture, the greater will be the proportion of the sphere of stimulated fluorescence which is collected and thus the greater will be the sensitivity of the lens. In practice, high numerical aperture lenses are often highly corrected and contain additional glass elements which may reduce the transmission efficiency of the lens. Similarly, the transmission characteristics of objectives at different wavelengths vary to a great degree and may be particularly reduced in the UV. For these reasons, selection of high numerical aperture objective lenses for FISH is best achieved by direct comparison of different examples on the microscope using fluorescent specimens of the type to be studied.

13.2.1 Electronic cameras

By placing an electronic camera at the image plane of the epifluorescence microscope, a digital image of the specimen can be obtained. A digital image consists of an array of picture elements (pixels) which contain binary coded measurements of the intensity or intensity and colour of the corresponding point in the optical image of the specimen. The advantage of digital over photographic recording is that the digital image can be processed and displayed directly by computer systems enabling enhancement and analysis of the image and convenient and rapid archiving and retrieval from digital storage media. Low-light level video cameras have been used for imaging on microscopes for over 20 years but recent advances in semiconductor device technology and the power of personal computers has allowed the development of highly sensitive and quantitative instruments. In recent years, several camera designs have been utilized for imaging fluorescence images. Of these, the most successful have been intensified video cameras and, more recently, cooled solid state detector arrays. It should be pointed out that the

development of solid-state detector arrays for fluorescence imaging is progressing at a rapid rate and camera systems are repeatedly superseded by superior designs.

13.2.2 Video cameras

Intensified video cameras, as the name suggests, comprise an image intensifier coupled to a video camera. The image intensifier serves to convert photons from the specimen into an electron image, to amplify the intensity of the image and to present it to the camera for conversion to a video output. Of the video camera types utilized for fluorescence imaging the silicon intensified camera (SIT) and its derivatives have proved most suitable. The SIT camera consists of an electrostatically focused image intensifier coupled to a silicon target camera within a single glass envelope. A more sensitive version of the SIT camera is the intensified silicon intensified target (ISIT) camera which utilizes an additional intensifier directly coupled to the photocathode of the SIT using fibre optics. Alternatively, the image intensifier can be employed as a unit separate from the video camera. In this configuration, the image intensifier typically uses a phosphor window as the output element which converts the electron image of the intensifier back into an optical image for viewing by the video camera. Lenses or fibre optic devices have been employed to couple the intensified image at the phosphor window onto the image plane of the video camera.

As these cameras operate at video rates of up to 30 frames per second, static images are acquired using a 'frame grabber' electronic circuit board which digitizes and stores a selected frame into computer memory or into a frame store. With appropriate electronic hardware, sequential images can be averaged into the frame store to improve the signal to noise ratio of the image or integrated to increase sensitivity at the expense of increased noise.

SIT and ISIT cameras have significant disadvantages for quantitative fluorescence imaging. Video cameras do not give a linear response between the video output and the intensity of the optical image and are susceptible to spatial distortion and thus are not optimal for advanced FISH techniques such as comparative genomic hybridization (CGH) (Chapter 8) or ratio labelling techniques (see Section 13.4.1) where the quantitative and spatial aspects of the imaged fluorescence is important. In addition, the sensitivity of video cameras varies across the field (an effect known as shading) and can be as great as a 30% difference in sensitivity from one part of the photocathode to another. Again, this feature,

which may be minimized using shading correction techniques, limits the suitability of such video cameras for CGH and ratio labelling.

13.2.3 CCD cameras

Video cameras are largely being replaced by solid-state detectors for FISH analysis due to the superior sensitivity and linearity of these modern devices. The most common solid state cameras used for FISH are based on charge-coupled device (CCD) technology. The photosensitive element of a CCD camera is a wafer of silicon onto which is applied a matrix of silicon dioxide and gate structures which form an array of photosensitive elements (Fig. 13.3). When a positive potential is applied to the gate of a CCD element, a depletion region is formed in the silicon base where photon-induced charge can be stored (the potential well). An image projected onto the CCD array produces a pattern of charge in proportion to the number of photons falling on each

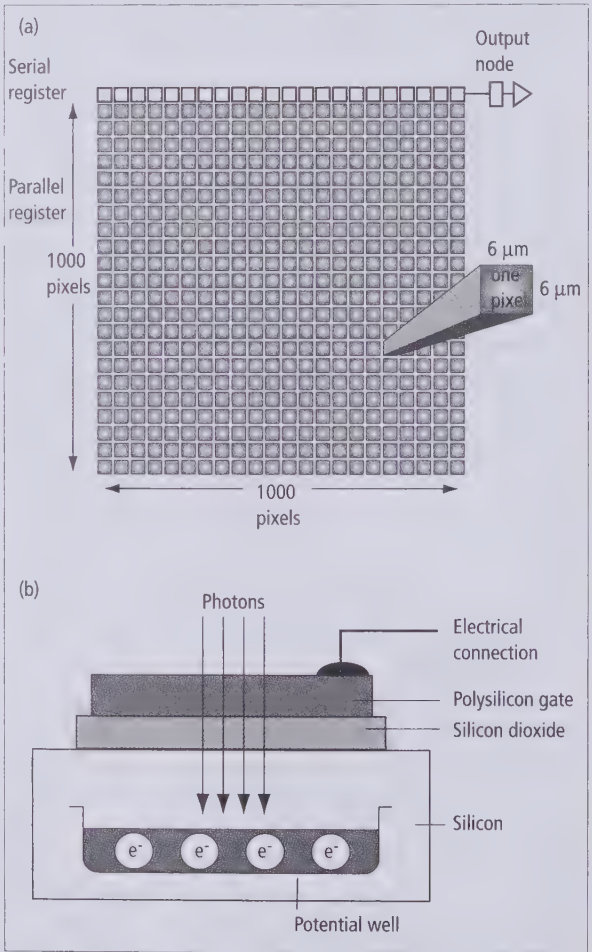


Fig. 13.3 Cooled CCD photodetector. (a) Arrangement of photodetectors in a CCD array. (b) Cross-section of a single CCD array element.

potential well. This pattern of charge stored in the photosensitive area (parallel array) is read out into computer memory or frame store by sequentially passing the charge row by row up the parallel array to a single row serial array for transfer to the output amplifier. Because charge is transferred in this way, CCDs require efficient charge transport from pixel to pixel on the chip. A scientific-grade CCD displays a charge transfer efficiency typically of 0.999998. A high charge transfer efficiency is of particular concern when imaging weak fluorescence signals which produce only a small charge in the potential well as charge loss during transfer would cause significant degradation of the image.

Integration of the image over time can be achieved directly on the CCD array without the need for an external frame store. A mechanical or solid state shutter is used to expose the CCD for imaging and block light from striking the array during image read-out. Increasing the length of the exposure increases the number of photons reaching each element of the array thus allowing direct integration within the dynamic range of the device. The highest quality slow scan CCD cameras display a noise-limited dynamic range of up to 105:1.

The sensitivity of CCD cameras is determined by the quantum efficiency of the device and by system noise. The quantum efficiency is a measure of the effectiveness of the device in converting photons into electronic charge. The quantum efficiency is always less than unity and varies with the wavelength of the light (Fig. 13.4). System noise is largely made up of photonic noise, dark current and preamplifier noise. Photonic noise (shot noise) and in particular the dark current and preamplifier noise combine to set the detection limit of the device. Photonic noise is due to the fundamental quantum nature of light and as such is unavoidable in imaging systems. Dark current is the accumulation of charge

on the CCD array with time and is thermally induced. Dark current is important when long on-chip integration times are employed for very weak fluorescence sources. Preamplifier noise is generated by the on-chip output amplifier and is important when photon-generated charge is small. Both long exposures and low photon-induced charge are often encountered in FISH experiments. Preamplifier noise is reduced by the use of the highest specification electronic components and optimal amplifier design while dark current can be reduced to levels of 0.1 electron per pixel per second by a combination of cooling the chip (typically to -25°C) and by the use of electronic biasing circuits.

Three types of CCDs are used in imaging systems, the full frame CCD, the frame transfer CCD and the interline transfer CCD. The full frame CCD employs a single photosensitive array for photon exposure, charge integration and charge transport. The frame transfer CCD has two parallel registers one of which is covered by an opaque mask and acts as a storage array. When the second register, the image array, is exposed, the electronic image is rapidly transferred to the storage array. While the storage array is read out, the image array is available for integration of the next image. In this way, the device can operate continuously at video rates. The interline transfer CCD has a parallel register divided so that alternate columns of pixels are masked and act as the storage register. The image integrated in the exposed area is transferred into the storage columns during image read-out. Interline transfer CCDs operate at video rates but exhibit reduced sensitivity because a large proportion of each array onto which the image falls is covered by the opaque mask. Both frame transfer and interline transfer CCD arrays can be operated without the need for a shutter to prevent light from reaching the device during image read-out.

The resolution of a CCD camera is determined by the physical arrangement of individual photosensitive elements in the array. CCDs with formats from $20\text{ }\mu\text{m}$ square pixels in a 512×512 array to $6\text{ }\mu\text{m}$ square pixels in a 4096×4096 array are now available. Full frame CCDs have no inactive regions as charge generated by a photon falling between pixels migrates to the nearest potential well. For the highest resolution, the CCD spatial sampling frequency should be at least twice the resolution of the microscope objective. The resolution of a $100\times$ immersion lens is approximately $0.2\text{ }\mu\text{m}$ which is efficiently matched to a 1000×1000 CCD array with $6\text{ }\mu\text{m}$ square pixels.

Some CCD cameras have additional features of subarray sampling and pixel binning. In subarray sampling, a region within an acquired image can be

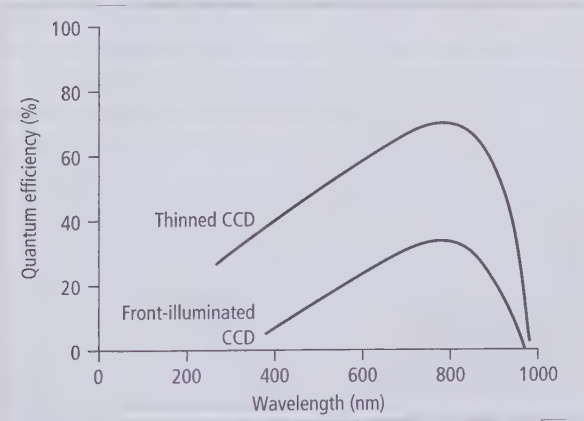


Fig. 13.4 Spectral characteristics of CCD arrays.

selected under computer software control so that subsequent images are acquired only from the corresponding area of the detector array. As fewer pixels need be transferred through the serial array, image transfer occurs at a much more rapid rate and automatic exposure algorithms react faster. Pixel binning allows the charge of adjacent pixels to be pooled reducing the spatial resolution of the display but increasing the sensitivity and dynamic range of the detector. For example a binning of 2×2 creates 'super pixels' in blocks of four which correlate with a single pixel on the computer image. These two features combine together for particularly effective analysis of FISH signals. An initial low-resolution image at a binning of, for example, 2×2 is acquired and displayed on the computer screen. A particular region of interest is then defined and a new image of this area acquired at high resolution (a binning of 1) in which each pixel of the detector corresponds to an individual pixel in the digital image.

Scientific grade slow scan CCD devices exhibit a very linear relationship between the intensity of the incident light and the digitized measurement over as much as five orders of magnitude and to within a few hundredths of a percent. On-chip light integration and high signal to noise ratios produced by cooling allow the imaging of the weakest fluorescent signals. Video rate CCD cameras do not display such a good linearity due to the use of high-speed, low-cost electronics and are less sensitive. For these reasons, the cooled, scientific grade slow scan CCD camera is currently the best (but most expensive) choice for FISH.

13.2.4 Colour CCD cameras

The CCD array is inherently a monochrome device having a relatively wide spectral response between 400 and 700 nm. However, by placing complementary colour filters in front of individual detector elements on the display array, each element can be made sensitive to one of three wavelength bands. The most common colour CCD cameras are of the interline transfer type operating at video rates with the filter matrix designed so either adjacent lines or individual elements are sensitive to different wavelengths. In this way a red, green and blue (RGB) output can be produced from the subarray of appropriated filtered detector elements to generate a true colour video image. Inevitably, the combination of interline transfer technology and the filter matrix compromises the spatial resolution of this type of camera. While uncooled colour CCD cameras can be used for some FISH applications, cooled colour CCD cameras provide enhanced sensitivity due to noise

reduction. Three-chip, cooled CCD colour cameras are now becoming available which do not suffer from the limitations of spatial resolution displayed by interline transfer cameras.

13.2.5 Imaging multiple fluorochromes with digital cameras

The filter block of the epifluorescence microscope allows the excitation wavelength of the arc lamp and emission spectra of the fluorochrome to be selected. Different fluorochromes demonstrate different excitation and emission spectra (see Appendix IV, Table IV.1). For example, DAPI demonstrates maximal excitation at 359 nm and emits at a maximum of 461 nm. DAPI is usually imaged using a filter block which selects an excitation wavelength around 365 nm and passes fluorescence above 420 nm (see Appendix IV, Tables IV.2 and IV.3 for tables of filter blocks). Similarly, fluorescein isothiocyanate (FITC) demonstrates maximal excitation at 495 nm and emits at a maximum of 519 nm. FITC is usually imaged using a filter block which selects an excitation waveband between 450 and 490 nm and passes fluorescence above 510 nm. Superimposition of DAPI and FITC images can be achieved in two exposures, one through each filter block. Unfortunately, the optical alignment of different filter blocks is rarely perfect and a lateral shift of the two images relative to each other occurs. This is a significant problem, particularly where the spatial relationship of one fluorescence signal to another is important such as in gene mapping.

This problem has been practically overcome by recent developments in optical filter technology. Filter blocks are now available which allow simultaneous excitation and detection of two, three or even four fluorochromes (Fig. 13.5). As a single filter block is used and not moved, image shift between fluorochromes due to filter block alignment is eliminated. With a colour CCD camera, a single exposure produces a colour image displaying all fluorochromes in registration. However, this multiple excitation approach cannot be used with a monochrome camera as the fluorescence from different fluorochromes would not be distinguished in the grey-scale image. For monochrome cameras, the multiple excitation filter is replaced by separate excitation filters, one for each fluorochrome, which are mounted in a motorized filter wheel placed in front of the lamphouse (Fig. 13.6). Under automatic computer control, the first filter is moved into the excitation light path and a monochrome image for the first fluorochrome is recorded. The second filter is then moved into place and an image for the

Fig. 13.5 Triple dichroic/bandpass filter block transmission characteristics.

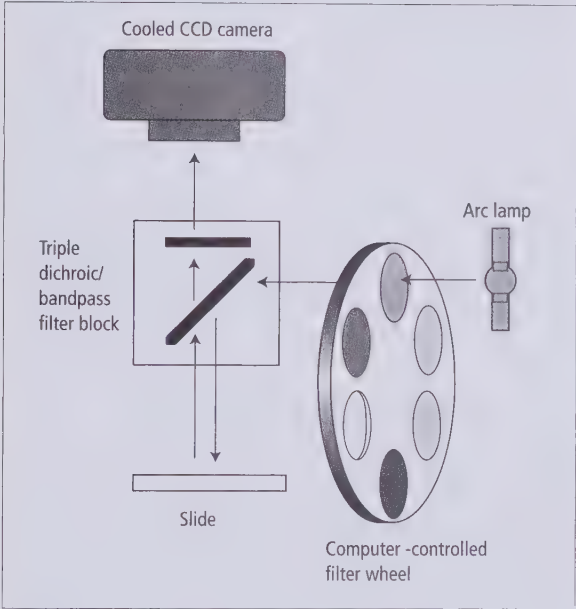
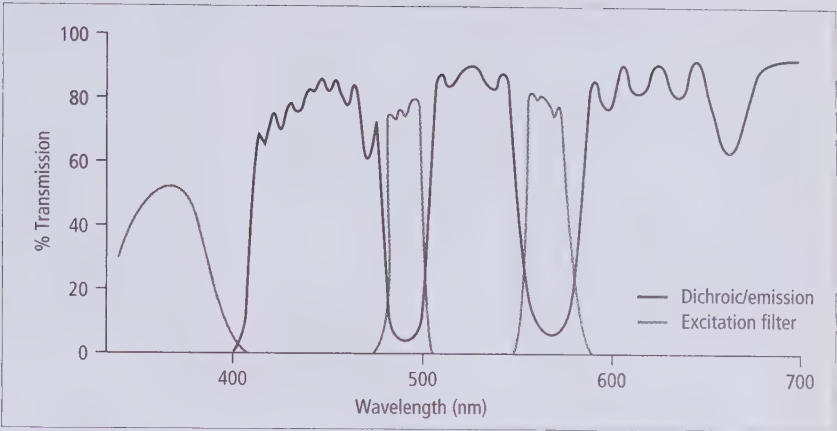


Fig. 13.6 Filter arrangement for a monochrome CCD camera.

second fluorochrome is recorded and so on until images have been acquired for all fluorochromes. The monochrome images are then merged into a colour image on the computer screen (see Plate 8).

The digitization of the image is another important process. Less expensive CCD arrays digitize typically at 8-bit resolution (256 grey levels) but cooled slow scan CCD arrays are now available which digitize at up to 16 bits (65 536 grey levels). The highest resolution colour images displayed on personal computers use 24 bits (millions of colours) made up of three 8-bit images, one for red, one for green and one for blue. Thus, if a coloured image is required by merging three monochrome fluorescence images into 24-bit colour from a camera which digitizes to more than 8 bits (e.g. 12 bits or 4096 grey levels), a normalization process is re-

quired to convert the 12 bits of information into the 8 bits to be incorporated into the coloured image. This normalization process can be a simple direct linear scaling but also allows for selection and transformation of the dynamic range to be converted. Thus, the part of the grey scale containing the information of interest can be selected for normalization to 8 bits thus retaining as much intensity resolution as possible.

One advantage of using a monochrome camera is that the exposure time can be adjusted for each fluorochrome. Thus, if one fluorochrome is bright, a short exposure is used while extended exposures can be used for weak signals. Some imaging systems allow for automatic adjustment of exposures by rapidly sampling the intensity of the image and adjusting the exposure to maximize the usage of the available dynamic range. This feature increases greatly the efficiency and speed of image acquisition.

13.3 Laser scanning microscopy

The *laser scanning microscope* utilizes a focused laser beam as the excitation source for imaging in the epifluorescence microscope (Fig. 13.7). The focused laser beam confines illumination and detection to a small spot which reduces the fluorescent flare from other regions. Fluorescence returning from the specimen passes back through the objective lens and is quantified by a photodetector. As the laser beam only illuminates a single point at a time, an image can only be built up by scanning the laser beam sequentially over the specimen. While this can be achieved by moving the specimen across the beam, most systems scan the beam across the specimen using electronically controlled galvanometer mirrors. The image is built up from the serial signal derived from the photomultiplier output as the laser beam moves from position to position. A series of

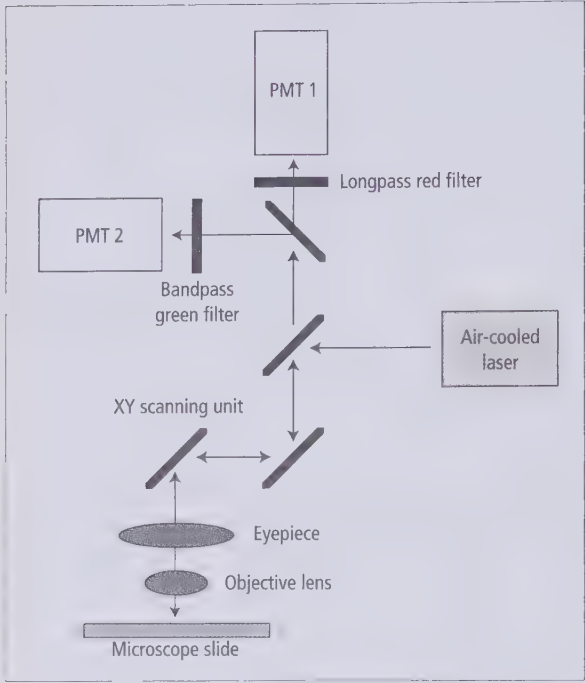


Fig. 13.7 Laser scanning microscope set-up. PMT, photomultiplier tube.

dichroic mirrors and interference filters allows specific fluorescence wavelengths to be detected at separate photomultipliers so that images from two or more fluorochromes can be collected simultaneously in registration.

The *confocal laser scanning microscope* utilizes an optical arrangement that rejects fluorescence from areas above and below the plane of focus and so allows optical sectioning of the specimen. The fluorescence returning from the specimen is focused on an aperture placed in front of the photodetector. Only fluorescence from the plane of focus passes efficiently through the aperture to the detector while fluorescence from above or below the plane of focus hits the aperture and is prevented from reaching the detector (Fig. 13.8). In this way, the confocal microscope produces a very narrow plane of focus which essentially generates an optical section of the

specimen, greatly enhancing the sharpness of fluorescent images.

Lasers can only produce light output at a small number of specific wavelengths defined by the energy levels of the electron orbits of the lasing medium (laser lines). The selection of fluorochromes whose excitation spectra match available laser lines is therefore restricted. Typical commercial confocal microscopes are supplied with air-cooled argon ion or mixed gas, argon/krypton lasers. With these lasers, DNA can be counterstained with propidium iodide (PI) while FISH signals can be detected with FITC. However, multicolour FISH techniques are more restricted, as the red PI fluorescent counterstain prevents the efficient use of red fluorochromes such as rhodamine, Texas Red or Cy3 which can be excited with these lasers. Counterstaining of DNA with DAPI or Hoechst requires excitation in the UV, necessitating a high-powered, water-cooled laser or a helium-cadmium laser, configured to operate together with the air-cooled argon ion or mixed gas laser. Such sophisticated and expensive confocal microscopes are not routinely used for FISH.

Fluorescence filter blocks on a typical confocal laser scanning microscope are listed in Appendix IV, Table IV.4.

13.4 Multiple probe detection

Many FISH applications, such as prenatal diagnosis of aneuploidy in fetal cells isolated from maternal blood or the ordering of DNA sequences, benefit from the simultaneous detection of multiple probes. Ideally, each probe would be labelled with a different fluorochrome so that each could be imaged separately. However, the excitation wavelengths available from arc lamps and lasers, and the limited number of excitation/emission wavebands that can be accommodated currently on dichroic filter blocks restricts fluorescence applications to the use of three or four fluorochromes. As it is normal to use one fluorochrome for counterstaining the DNA, probe labelling is restricted to two or three fluorochromes and thus the direct detection of only two or three

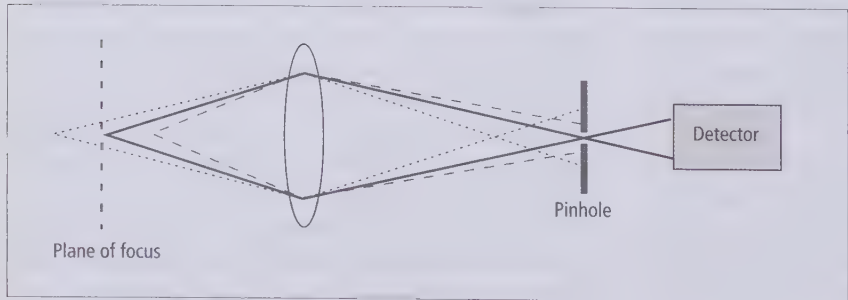


Fig. 13.8 Principle of optical sectioning in the confocal microscope.

different probes. Nederlof *et al.* [16] suggested that labelling of a probe with two fluorochromes simultaneously would generate an intermediate colour, allowing that probe to be distinguished from other probes labelled only with a single fluorochrome. In this combinatorial scheme, two fluorochromes (e.g. red and green) would allow the detection of three probes (red only, green only, red and green together giving yellow; see Plate 8) whereas three fluorochromes would allow the detection of seven probes (see also Chapter 10, Section 10.1).

13.4.1 Ratio labelling

This scheme can be extended by not only labelling probes with equal proportions of each fluorochrome but also with differing proportions of the fluorochromes in ratio labelling schemes. Thus a probe that is labelled with 70% red and 30% green would be distinguishable from a probe labelled with 50% red and 50% green. Using direct labelled chromosome paints, it has been found that up to eight different chromosomes can be distinguished with two fluorochromes and 15 different chromosomes can be distinguished with three fluorochromes. Indirect detection systems introduce greater variation into the system such that the number of chromosomes that can be distinguished in this way is reduced. For example, only five probes can be reliably distinguished with two fluorochromes using indirect detection of the hybridization signals. Probes can be labelled directly with a proportional mixture of modified nucleotides by nick translation or the polymerase chain reaction or alternatively each probe can be labelled separately with each label or hapten and then mixed in proportion before hybridization. The latter scheme

allows for adjustment of the proportion of the labels in subsequent hybridizations and as such is more versatile than direct ratio labelling.

For smaller probes, such as cosmids, it can often be difficult to determine the colour visually according to the ratio of labels used as only a small area of the image and few pixels are available for assessment. Image processing of the digital image allows a more objective method for determining the ratio of probe utilizing the measured intensities of each fluorochrome in a defined region, as a moving average or on a pixel by pixel basis. Specific fluorochrome ratio ranges can be set using the image analysis software which define each probe and to aid visualization further, the signal colour in the image can be replaced by a pseudocolour of choice (see Plate 9).

13.5 Fluorescent chromosome band enhancement

Image processing of digital images allows specific enhancement of the information contained in the image. For metaphase chromosome analysis, it is usual to relate results to the chromosome banding pattern. *Fluorescence banding techniques* [17] that are consistent with the simultaneous collection of FISH signals produce banding patterns of low contrast. Processing of the image using high-pass spatial filters ('Mexican hat' filters) enhances the difference in contrast between adjacent regions and is particularly useful for highlighting chromosome bands from DAPI or DAPI/PI counterstained chromosomes. The filter is applied to the image pixel by pixel, where each pixel value is adjusted to the sum of the surrounding pixel values multiplied by the factor defined by the linear filter matrix (Fig. 13.9).

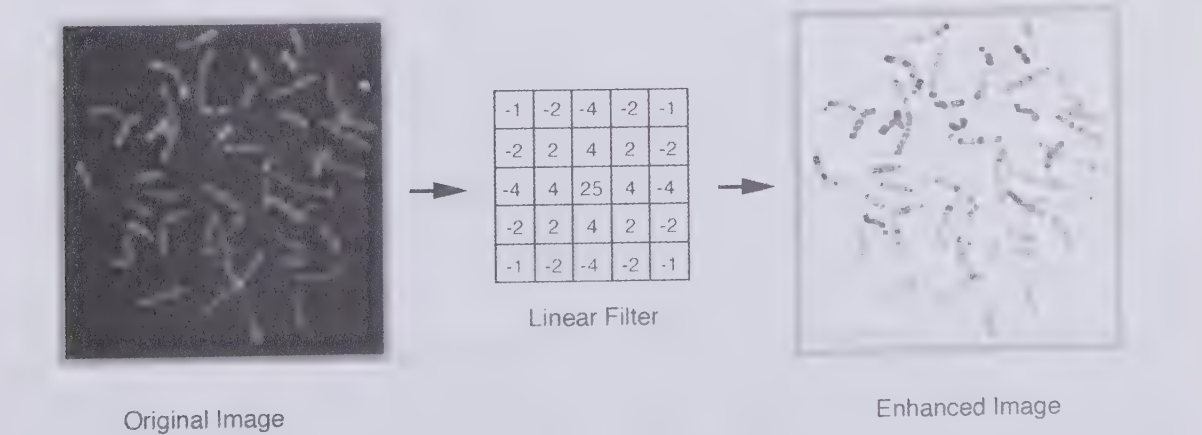


Fig. 13.9 Enhancement of fluorescent bands by image processing with a linear filter. The original image was

processed with the linear filter, colour reversed and contrast adjusted to produce the enhanced image.

13.6 Obtaining hardcopy output

An important part of the digital acquisition and processing of FISH images is the production of high-quality, hardcopy output in the form of transparencies and prints. Colour prints of digital images can be produced by connecting a video printer directly to the computer screen to generate output. These devices are restricted to the resolution of the screen which is often less than the resolution of the acquired image. A better alternative, which is not restricted by the resolution of the screen, is to use a colour printer attached to the personal computer. A wide range of colour printers are now available for the production of coloured prints. The cheapest and lowest quality output is available from colour ink-jet printers. These printers produce a matt coloured output onto plain paper but often generate a banded effect in large blocks of a single colour (FISH images often contain large areas of black). Wax transfer printers produce higher quality images but tend to be restricted in the number of colours that can be represented. Dye sublimation printers currently provide the best choice for the printing of FISH images. Output is in continuous tone at full 24-bit colour onto special high gloss paper and prints are of a quality adequate for direct publication. Unfortunately, these printers are at present expensive to buy and to run.

Similarly, production of high quality transparencies from digital FISH images requires a 4000-line film recorder which is also expensive to buy. However, transparencies are usually required less frequently than prints and are probably most cost effectively produced by taking an image file to a computer film bureau or, in academic institutions and hospitals, by using the services of an audio visual aids department.

A major problem for colour reproduction is that the colour characteristics of the computer monitor are invariably different from the capabilities of colour output devices. This colour-matching problem results in the colour print or transparency often being a disappointing representation of the image on the computer screen. This is particularly true for the blue image of DAPI counterstain. When viewed using the microscope, DAPI fluorescence is light blue/cyan in colour but using a monochrome camera the image is displayed as deep blue. When printed, the blue image reproduces darkly with little contrast to the black background. However, this blue image can be recoloured to cyan by adding a proportion of the blue image to the green image before colour merging. The combination of counterstain signal in both the blue and green planes of the

merged image produce a cyan colour closer to the true colour of DAPI. This is achieved by splitting the colour planes of the image and applying simple image arithmetic to the blue and green images before remerging of the red, blue and green images. This series of operations can usually be automated using the macro language feature of most scientific image analysis software. Alternatively, the same effect can be achieved using a specific colour affine matrix transformation feature in packages such as IPLab Spectrum or the lightening function of the Hue and Saturation tool of software such as Adobe Photoshop (see Plate 7).

13.7 Image data storage

Digital images use a large amount of computer memory. A 1300×1000 image in 24-bit colour generates a file which is of the order of 5 megabytes (Mbytes) in size when stored from the image-processing software. It is therefore necessary to have not only a large amount of system RAM in the personal computer (e.g. 32 Mbytes) but also a suitable storage medium. While large hard disk drives are now available with greater than 1 gigabyte of storage, these will eventually become full and a removable storage medium will be required for data archiving. It is clear from the potentially large size of images that floppy disks which currently store up to 1.4 Mbytes are totally inadequate for this purpose, even if data compression algorithms are used. Of the removable storage devices currently available, optical disks provide a good compromise between speed of operation, media costs and storage capacity. Currently, typical 3.5-inch optical disks have capacities of up to 128 Mbytes, while 5.25-inch disks can store over 1 gigabyte. These disks can be used for archiving only as WORM drives (write once read many) or as erasable media. For archiving of data where rapid access is less important, digital audio tape (DAT) drives are most appropriate. DAT tapes will store 2, 8 or even 16 gigabytes of data with rapid hard-wired compression and have storage costs of as little as £1 per gigabyte. An optimal system would use optical storage for medium-term storage of data with long-term archiving and backup onto DAT tape.

Acknowledgements

I would like to acknowledge the support of my colleagues in the Human Molecular Genetics Group, Department of Pathology, University of Cambridge and in the East Anglian Cytogenetics

Service, Addenbrooke's Hospital, Cambridge, and in particular Yun-ling Zheng who provided the original data for Plates 8 and 9.

References

1 Viegas-Pequignot, E., Dutrillaux, B., Magdelanat, H. & Coppey-Moisan, M. (1989) Mapping of single-copy DNA sequences on human chromosomes by in situ hybridization with biotinylated probes: enhancement of detection sensitivity by intensified-fluorescence digital-imaging microscopy. *Proc. Natl Acad. Sci. USA* **86**, 582–586.

2 Trask, B., Pinkel, D. & van den Engh, G. (1989) The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. *Genomics* **5**, 710–717.

3 Lichter, P., Boyle, A.L., Cremer, T. & Ward, D.S. (1991) Analysis of genes and chromosomes by nonisotopic in situ hybridization. *Genet. Anal. Tech. Appl.* **8**, 24–35.

4 Lawrence, J.B., Singer, R.H. & McNeil, J.A. (1990) Interphase and metaphase resolution of different distances within the human dystrophin gene. *Science* **249**, 928–932.

5 Wiegant, J., Kalle, W., Mullenders, L. *et al.* (1992) High resolution in situ hybridization using DNA halo preparations. *Hum. Mol. Genet.* **1**, 587–591.

6 Parra, I. & Windle, B. (1993) High resolution visual mapping of stretched DNA by fluorescent hybridization. *Nature Genet.* **5**, 17–21.

7 Fidlerova, H., Senger, G., Kost, M. *et al.* (1994) Two simple procedures for releasing chromatin from routinely fixed cells for fluorescence in situ hybridization. *Cytogenet. Cell. Genet.* **65**, 203–205.

8 Carter, N.P. (1994) Cytogenetic analysis by chromosome painting. *Cytometry* **18**, 2–10.

9 Grifo, J.A., Boyle, A., Tang, Y.X. & Ward, D.C. (1992) Preimplantation genetic diagnosis: in situ hybridization as a tool for analysis. *Arch. Path. Lab. Med.* **116**, 393–397.

10 Griffin, D.K., Handyside, A.H., Penketh, R.J. *et al.* (1991) Fluorescent in situ hybridization to interphase nuclei of human preimplantation embryos with X and Y chromosome specific probes. *Hum. Reprod.* **6**, 101–105.

11 Habeebu, S.S., Spathas, D.H. & Ferguson-Smith, M.A. (1990) Nonradioactive in situ hybridization of DNA probes to chromosomes and nuclei. A comparison of techniques. *Mol. Biol. Med.* **7**, 423–435.

12 Ried, T., Landes, G., Dackowski, W. *et al.* (1992) Multicolor fluorescence in situ hybridization for the simultaneous detection of probe sets for chromosomes 13 (18), 21, X and Y in uncultured amniotic fluid cells. *Hum. Mol. Genet.* **1**, 307–313.

13 Dauwerse, J.G., Wiegant, J., Raap, A.K. *et al.* (1992) Multiple colors by fluorescence in situ hybridization using ratio-labelled DNA probes create a molecular karyotype. *Hum. Mol. Genet.* **1**, 593–598.

14 Nederlof, P.M., van der Flier, S., Wiegant, J. *et al.* (1990) Multiple fluorescence in situ hybridization. *Cytometry* **11**, 126–131.

15 Wiegant, J., Wiesmeijer, C.C., Hoovers, J.M. *et al.* (1993) Multiple and sensitive fluorescence in situ hybridization with rhodamine-, fluorescein- and coumarin-labeled DNAs. *Cytogenet. Cell. Genet.* **63**, 73–76.

16 Nederlof, P.M., Robinson, D., Abuknesha, R. *et al.* (1989) Three-color fluorescence in situ hybridization for the simultaneous detection of multiple nucleic acid sequences. *Cytometry* **10**, 20–27.

17 Verma, R.S. & Babu, A. (1989) *Human Chromosomes, Manual of Basic Techniques* (Pergamon Press, Oxford).

18 Van Dilla, M.A., Dean, P.N., Laerum, O.D. & Melamed, M.R.(eds). (1985) *Flow Cytometry: Instrumentation and Data Analysis* (Academic Press, London).

Section 3

**Physical mapping 2:
long-range mapping and
gene isolation**

Section 3

Introduction

Anna-Maria Frischauf

ICRF, Lincoln's Inn Fields, London WC2A 3PX, UK

New techniques have led to an ever-increasing rate of physical and genetic mapping of the genomes of different species. The chapters in this section describe physical mapping and gene isolation techniques. Most methods are applicable to all higher organisms. There are, however, big differences between organisms in the resources and information that are already in the public domain. The human genome has been the focus of the greatest effort, and physical mapping and gene identification have progressed the furthest. In the immediate future, an integrated, high-resolution genetic and physical map of the human genome, complete with the position of the majority of transcribed, polyadenylated sequences (expressed sequence tags, ESTs) will be available. It is unlikely that other animal genomes will reach the same degree of coverage in the near future, but it will be possible to transfer subsets of the available human data into the physical and genetic framework map of other mammalian species, since the arrangement and order of large groups of genes are usually conserved between many different species. Chapter 37 and Appendix V list web sites and distribution centres for information and resources on human and other genomes.

The methods described in this section can be applied to the initial physical mapping of species on

which very little information exists. Alternatively, they can be used to complement and refine existing data on human or other well-researched genomes. Comprehensive physical maps at a wide range of different resolutions can be obtained by somatic cell genetics as described in Chapter 14 (A. Schafer & C. Farr). In particular, irradiation fusion hybrids, which contain irradiation-induced chromosomal fragments from the required species on a hamster or mouse background, can be used to order markers over wide physical distances in a manner not too dissimilar from genetic mapping. Radiation hybrid mapping is useful in that it does not require markers that are polymorphic within the species of interest, it is only necessary to be able to distinguish the marker from any homologues in the mouse or hamster background.

Most physical mapping projects then aim at obtaining a cloned representation of a genomic region. The construction of the required yeast artificial chromosome, P1 and cosmid libraries is described in Chapter 15 (S. Meier-Ewert, L. Schalkwyk, F. Francis & H. Lehrach). If the region in question is part of the human genome, these resources are already available (as are P1 artificial chromosome (PAC) and bacterial artificial chromosome (BAC) libraries) and can be obtained as filters for screening or as pools for PCR from the Reference Library Data

Base, the UK Human Genome Mapping Project Resource Centre and other publicly funded and commercial sources (see Chapter 37 and Appendix V). Many of these resources are also available for the mouse (see Chapter 26). For other species, the situation is variable and will have to be explored. The large range of insert sizes obtainable with different cloning vectors can be exploited to combine fast coverage of a region with the convenience and high resolution afforded by small insert vectors. A few markers at long intervals will frequently suffice to cover a region with yeast artificial chromosomes. These can then be converted to cosmids either directly, by subcloning, or indirectly by screening a chromosome-specific library (now available for most human chromosomes). The ordering of a large number of clones covering a region into a contig is a problem that requires a good strategy to minimize the experimental work and take into account all available information to construct correct maps. Chapter 16 (R. Mott, A. Grigoriev & H. Lehrach) describes various strategies for doing so. These approaches can be used for regions of around a megabase to whole mammalian chromosomes.

The identification of transcripts within genomic DNA is usually the next goal in the search for a

phenotype associated with a region of genomic DNA. Many approaches are available; the two currently most popular complementary strategies—exon trapping and cDNA selection—are described in Chapter 17 (M. North *et al.*). As the mapping of human ESTs proceeds, it will be possible to test ESTs that have been mapped to the approximate region directly for their presence in the clone contig DNA. As sequencing speeds improve and information accumulates, more genes will also be able to be identified by computer predictions based on the sequence of genomic DNA.

Chapter 18 (D. Simmons) describes a special case of gene searching which is not based on genetic, but on cell biological and functional, information. Such techniques are the most efficient route to genes for which such handles exist; but at present the function of the majority of genes is unknown and for most it is only possible to confidently predict function on the basis of sequence similarity to known proteins. The final step in cloning a disease gene is to verify whether a gene that has been cloned using the methods described above is, in fact, responsible for the disease. Mutation analysis has to be performed on patient DNA and Chapter 19 focuses on an efficient method to do so: denaturing gradient gel electrophoresis (R. van der Lijst and R. Fodde).

Chapter 14

Somatic cell hybrid approaches to genome mapping

Alan J. Schafer & Christine J. Farr

Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK

14.1	Introduction, 323		
14.2	Hybrid biology and types of hybrids, 324		
14.3	Mapping hybrid phenotypes, 324		
14.4	Mapping cloned DNA, 325		
14.5	General considerations, 325		
14.5.1	Cell culture, 325		
14.5.2	Mycoplasma contamination, 326		
14.5.3	Fusogens, 326		
	Troubleshooting, 327		
	PEG toxicity, 327		
14.6	Source of cells, 327		
14.7	Chromosome segregation, 328		
14.8	Selection, 328		
14.8.1	Endogenous selection genes, 332		
14.8.1.1	HAT selection, 332		
14.8.2	Biochemical selections for other endogenous markers, 333		
14.8.3	Other selection schemes for endogenous markers, 333		
14.8.4	Hybrids made from auxotrophic mutants, 333		
14.9	Insertion of dominant selectable markers into the mammalian genome, 334		
14.9.1	Dominant selectable markers, 334		
14.9.1.1	Transfer of dominant selection genes into mammalian cells, 334		
14.9.1.2	Eukaryotic expression vectors, 334		
14.9.2	Selection schemes, 335		
14.9.2.1	Positive selection, 335		
14.9.2.2	Negative selection schemes, 338		
14.9.2.3	Positive–negative bidirectional selection systems, 338		
14.10	Whole-cell fusion, 338		
14.11	Microcell-mediated chromosome transfer, 339		
14.11.1	Micronucleation, 340		
14.11.2	Isolation and enucleation of microcells, 341		
14.11.3	Filtration, 341		
14.11.4	Fusion, 342		
14.12	Irradiation and fusion gene transfer (radiation hybrids), 342		
14.12.1	Selection, 342		
14.12.2	Radiation dose, 343		
14.12.3	Irradiation and fusion, 343		
14.12.4	Analysis of radiation hybrids, 343		
14.12.5	Making the map, 344		
14.13	Characterization of somatic cell hybrids, 344		
14.13.1	Marker analysis, 344		
14.13.1.1	Interspersed repetitive sequence-PCR, 344		

Protocol 68

Preparation of HAT medium, 346

Protocol 69

Whole-cell fusion of mammalian cells, 347

Troubleshooting, 349

Too few clones, 349

Hybrids with no donor chromosomes, 349

Protocol 70

Optimization of micronucleation, 350

Protocol 71

Enucleation from plastic bullets, 351

Protocol 72

Percoll gradient enucleation: an alternative protocol for enucleation, 354

Protocol 73

Filtration of microcells, 354

Protocol 74

Fusion of microcells to whole recipient cells, 355

Troubleshooting, 356

Low hybrid yield, 356

Problems with micronucleation, 357

Problems with enucleation, 357

Problems with fusion, 357

Hybrids with no donor chromosomes, 357

Hybrids with too many donor chromosomes, 357

Protocol 75

Production of radiation hybrids, 358

References, 358

14.1 Introduction

Since its development over 30 years ago, the somatic cell hybrid, the basic tool of somatic cell genetics, has acquired a central role in gene mapping in a wide variety of organisms. In general, somatic cell hybrids are used for two distinct, but non-exclusive, types of mapping project:

- 1 the mapping of hybrid phenotypes; and
- 2 the mapping of cloned DNA sequences.

In *phenotype mapping*, a characteristic conferred on the hybrid by one of the parental cells is associated with that parent’s genetic contribution to the hybrid. For mapping cloned DNA sequences, retention of a marker in a hybrid is correlated with a hybrid karyotype or, alternatively, coretention of multiple markers from one parental genome is used to establish synteny and linkage between the markers.

These two types of mapping can utilize similarly derived hybrids, but the types of parental cells used and most appropriate type of hybrid will depend upon the objective of the experiment. This chapter considers general strategies for mapping using somatic cell hybrids and describes methods for producing hybrids with a reduced genetic complexity of donor DNA (Table 14.1). The protocols presented here have been widely used in human–rodent and rodent–rodent fusions, but the methods have been applied to cells from many mammalian sources, and identical protocols have been used for avian cells. Fusion protocols have been developed for cells from other eukaryotic organisms, including fungi [1] and plants [2], but these methods tend to be specialized and are outside the scope of this chapter.

Somatic cell hybrids can be used to:

- map human and other mammalian chromosomes by analysis of hybrid phenotypes and correlation with karyotype
- map cloned DNA sequences

They can be used in positional cloning strategies to:

- map a disease gene to a chromosome
- supply additional markers
- localize markers within the disease locus region
- partition deleted and rearranged chromosomes that are associated with the phenotype

Irradiation and fusion gene transfer can be used to:

- isolate hybrids containing small defined regions of specific chromosomes
- construct whole chromosome (or whole genome) maps, commonly known as radiation hybrid mapping (or radiation fusion hybrid mapping)

Microcell-mediated chromosome transfer can be used for:

- transferring individual chromosomes
- chromosomal assignment and subchromosomal mapping of DNA sequences
- cloning DNA markers and expressed genes
- mapping and identifying loci by a phenotype conferred upon recipient cells: for example, differentiation, DNA repair, senescence, tumorigenesis and apoptosis

Applications box 14.1

Table 14.1 Interspecific hybrids and their applications.

	Whole-cell hybrids	Microcell hybrids	Radiation hybrids
Application	Partition donor chromosomes in hybrids	Transfer individual or few donor chromosomes into recipient cell	Mapping of DNA sequences
Construction	Fuse donor and recipient cells	Micronucleate donor cells Isolate microcells by enucleation Fuse to recipient cells	Irradiate donor cells and fuse to recipient cells
Advantages	Simple to make High-frequency fusion	Low-complexity hybrids Specific chromosome transfer	Simple to make Excellent mapping reagents
Disadvantages	Hybrids usually complex with multiple donor chromosomes	Low-efficiency fusion Technically difficult	Large hybrid panel Instability of hybrids requires large-scale DNA isolation from a single time point

14.2 Hybrid biology and types of hybrids

The spontaneous fusion of mammalian cells *in vitro* and the culture of the first hybrid cells were demonstrated conclusively by Barski *et al.* [3] in 1960, but the low frequency of spontaneous fusion initially limited the use of somatic cell hybrids as an experimental tool. Their widespread use became possible with the discovery of biological and chemical *fusogens* that promote cell fusion, and the development of selection systems for hybrid cells (see Section 14.8). Exposure of a mixture of whole cells to a fusogen induces fusion of the cytoplasmic membranes to form *heterokaryons*, which contain both the parental cell nuclei in a common cytoplasm surrounded by a plasma membrane.

Fusion is not limited to two cells, for several may fuse simultaneously or sequentially to yield multinucleate cells containing varying numbers of nuclei. However, only hybrids produced by the fusion of a small number of nuclei can propagate successfully and most experiments have made use of fusions between only two cells. When chromosome replication and nuclear mitosis take place in these cells, those nuclei that enter mitosis together are usually reconstituted as a single unit containing the chromosomes of both parental genomes. Such cells are described as *synkaryons* or simply as somatic cell hybrids. Hybrids made by the fusion of intact parental cells are *whole-cell hybrids*. The amount of DNA retained from one of the parents can be much reduced in the hybrid and will vary depending on the species origin and type of cells fused, as well as on the method used to generate the hybrids. Interspecific hybrids have two great advantages over intraspecific hybrids for mapping.

1 One set of parental chromosomes (those of the 'recipient' cell) is preferentially maintained as a background, while the other parental chromosomes (from the 'donor' cell) tend to be eliminated, resulting in the loss of chromosomes of this species and thus a reduction in the complexity of the donor genetic material.

2 It is possible to determine the genetic contribution of each parental cell line to the hybrid by karyotypic or marker analysis.

Selection for a marker present in the genome of the donor and not present in the recipient cell ensures that the donor chromosome (or a portion of the chromosome) containing the marker is retained in the hybrid. However, some unselected donor chromosomes may also be fortuitously retained (with varying stability) as independent chromosomes, or as translocations or insertions onto

recipient chromosomes. In interspecific hybrids, the direction of chromosome loss depends primarily on the particular species combination, but can be affected by other factors, such as cell type (see Section 14.7).

Other types of hybrid are produced by modified whole-cell fusion protocols. *Microcell hybrids* (see Sections 14.3 and 14.11) are derived from the fusion of micronuclei (subnuclear packets containing a subset of the donor genomic chromosomes) with intact recipient cells (see Protocols 70–74). The reduction in donor DNA in the microcells results in hybrids of lower complexity than whole-cell hybrids.

Radiation hybrids are generated by the technique of irradiation and fusion gene transfer (IFGT) (see Section 14.12), which involves the whole-cell fusion of an irradiated donor cell with a nonirradiated recipient cell line (see Protocol 75). Irradiation breaks the donor cell chromosomes, so the resultant hybrids contain many fragments of donor chromosomes, most of which are retained independently of selection.

14.3 Mapping hybrid phenotypes

Early cell fusion experiments focused on the phenotypic properties of the hybrids and the ways in which they differed from those of their parental cell lines. Many phenotypes have been studied, and loci have been identified that modulate gene expression [4], induce cellular senescence [5], suppress tumorigenicity [6] and metastasis [7], and complement cellular defects, such as those involved in the human DNA repair disorders xeroderma pigmentosum [8] and ataxia telangiectasia [9].

Initial experiments to map a locus involved in a phenotype are usually whole-cell fusions to determine if the phenotype of one of the parental cells can be conferred on the hybrid. For example, a non-tumorigenic cell line can be fused with a tumorigenic one and the hybrids tested for tumorigenicity. A collection of whole-cell hybrids with different subsets of donor chromosomes is then used to identify candidate chromosomes mediating the phenotype by correlation of the hybrid karyotype with the presence or absence of the phenotype. Subsequent experiments involve fusions with donor cells contributing less- or better-defined pieces of DNA to the hybrids.

The transfer of normal chromosomes via microcell-mediated chromosome transfer (MMCT) (see Section 14.11) has been instrumental in such studies. Individual candidate chromosomes can be transferred and the hybrids tested for the phenotype. Alternatively, chromosomes can be randomly marked by transfection of cells with a dominant

selectable gene prior to transfer. These cells are used as donors in MMCT and a genotype/phenotype correlation is established. Identification of the donor chromosomes in the hybrids must be unequivocal, so interspecific hybrids are usually produced. A proportion of microcell hybrids will contain deletions of the donor chromosome and these deletion hybrids can be used for further localization of the locus. Another course is to irradiate donor cells before transfer, in order to fragment the chromosomes. An important caveat to phenotype mapping is that hybrid phenotypes of polygenic origin may manifest themselves in whole-cell hybrids, but may not be evident in hybrids of reduced genetic complexity.

Mapping a phenotype can be difficult if there is no selectable or readily discernible feature conferred on the hybrids. A large number of hybrids may need to be tested before one is found that exhibits the phenotype or it becomes clear that the phenotype does not occur. Another complicating factor is the possibility of undetected donor DNA being responsible for the phenotype in some of the hybrids, which can easily misdirect mapping [10]. Secondary transfers of chromosomes can help to exclude this possibility. Hybrid genotypes should be as fully characterized as possible and multiple hybrids need to be used to map the locus.

14.4 Mapping cloned DNA

Somatic cell hybrids have been widely exploited to map genes and markers in mammalian genomes, in particular the human genome. In the early 1970s several groups initiated the systematic application of somatic cell hybrids to the mapping of human chromosomes [11]. Initially, genes were mapped by isozyme analysis, but with the advent of recombinant DNA techniques, this approach is now uncommon. In the intervening years, many interspecific hybrids with well-defined subsets of mammalian genomes have been made and chromosomal mapping panels for the human genome are now widely available. Two panels that each include all human chromosomes are available from the NIGMS Human Genetic Mutant Cell Repository (Camden, NJ, USA; see Appendix V for address). One set consists of hybrids with a reduced number of human chromosomes and the second consists of hybrids with only one human chromosome (monochromosomal hybrids). These sets of hybrids can be used to map a DNA marker to a specific chromosome by PCR or Southern analysis of the hybrid collection. An advantage of working with hybrids retaining multiple chromosomes is the

internal redundancy, which may strengthen chromosomal assignment.

A more recent application of somatic cell hybrids has been in the positional cloning of human disease genes. In such experiments, hybrids have been used to map the gene to a chromosome [12, 13], supply additional markers [14], localize the markers within the disease locus region [15], and partition deleted and rearranged chromosomes that are associated with the phenotype for mapping and cloning experiments [16].

Interspecific somatic cell hybrids are used in several ways to map cloned DNA sequences.

1 Entire unrearranged chromosomes can be partitioned in hybrids by whole-cell fusion or by MMCT. The donor DNA contribution is established by karyotype and marker identification, and the hybrids are used to map sequences to the retained donor chromosomes. Chromosomal deletions occurring as a result of fusion can serve to sublocalize sequences, but these chromosomes often contain other rearrangements that complicate mapping.

2 Chromosomes with naturally occurring rearrangements can be partitioned in hybrids. Hybrids constructed from cells containing a chromosome with a cytologically detectable break associated with a phenotype provide a resource for mapping of markers relative to the breakpoint. A collection of hybrids with rearranged chromosomes from different patients can be used to map a disease locus by defining shared regions of overlap or deletion.

3 DNA markers can be mapped relative to one another using IFGT. The likelihood of two markers being separated by a radiation-induced break in the DNA is a function of physical distance, so that markers closer together have a higher probability of coretention in any given hybrid than markers further apart. Examination of marker coretention frequencies in a panel of radiation hybrids is used to generate a radiation map of the donor genome. Radiation hybrid (RH) mapping has proved extremely effective for mapping markers. Producing a radiation panel is laborious, but for any large-scale mapping effort it is the method of choice. Furthermore, radiation hybrids may retain small enough quantities of human DNA to facilitate cloning of the gene, or closely linked markers, directly from the hybrid.

14.5 General considerations

14.5.1 Cell culture

Cell lines used for fusion experiments should be managed under conditions of optimal growth as

subconfluent, exponentially growing cultures are best for fusion. The medium used for cultivation of cells should be the one most favourable to growth, as the type of medium does not affect cell fusion. If the two cell lines to be fused differ greatly in their growth medium, plate them into the recipient cell growth medium following cell fusion.

Penicillin and streptomycin are commonly added to tissue culture media to prevent bacterial infection of cultures. While this may diminish the possibility of bacterial contamination, it will also mask poor sterile technique which can result in the introduction of other organisms unaffected by these antibiotics, such as yeast and mycoplasma (see Section 14.5.2). Cells should therefore routinely be grown without antibiotics, including them only during handling when the chance of contamination is increased (e.g. microcell fusion). Antimycotics, such as mycostatin and amphotericin B (fungizone), should be reserved for use in treating contamination of a valuable stock. Ideally, any contaminated culture should be discarded. As a general rule, use the best-quality chemicals and reagents available, including 'tissue culture grade' reagents available from several distributors.

In most cases, hybridized cells are plated at densities such that clonal populations can be easily isolated. Plating efficiencies should therefore be determined for each recipient cell line. Increasing serum concentration (i.e. up to 20%) can dramatically improve cell viability, growth rate and plating efficiency, and different serum batches can vary widely in their ability to support cell growth. Optimize serum concentration for each cell line and test each new serum batch at various concentrations for maximum plating efficiency. Cell lines with poor plating efficiencies should be avoided or fusion protocols should be adjusted accordingly to compensate.

Rough handling of postfusion cultures can dislodge cells from a hybrid clone. The cells may reattach and appear as independent clones but will be duplicates, increasing the number of hybrids to be analysed. Various methods can be used to isolate (pick) clonal populations of cells. Cloning rings are commonly used, but require clear access to the cell colony, achieved either by cutting the top off a flask with a hot scalpel or soldering iron, or by plating cells into petri plates (which run a higher risk of contamination than flasks). Clones can also be harvested with a sterile pipetman tip or, if the clones are well spaced, they can be harvested with cotton swabs dipped in trypsin. An alternative method for picking clones from flasks is to use sterile, plugged Pasteur pipettes that have had their ends bent at a

45–90° angle \approx 1 cm from the tip. Using a rubber bulb, a small amount of medium is drawn into the pipette. The pipette is then inserted through the neck of the flask (having first removed the medium) and the clone is scraped free with the pipette tip. The medium in the pipette is used to carefully rinse the area, drawn back into the pipette, and transferred into 25-cm² flask. With practice, this is an easy and efficient way to harvest large numbers of clones.

The hazards associated with working with cells in culture are not fully known. Any primate-derived cell line might contain viruses transmissible to humans. An undetected latent virus could be activated, so even common cell lines should be suspect and handled appropriately. Cells used in culture and for fusions should never be derived from a person who will be working with or around the cells. A transformed or hybrid cell line may develop tumorigenic properties, while retaining the 'self' phenotype of the donor of the cells, posing an unknown and unnecessary risk.

14.5.2 Mycoplasma contamination

Mycoplasma are prokaryotic microorganisms of the order Mycoplasmatales that can infect tissue culture cells. These organisms lack a cell wall and are resistant to antibiotics such as penicillin and streptomycin. Mycoplasma infection of tissue culture cells can cause changes in genotype [17], morphology [18], growth rate [19], metabolism [20] and membrane structure [21], and result in poor hybrid yields. Contamination is not necessarily obvious in the manner of bacterial or fungal contamination, which usually cause turbidity or pH change of the medium. It is therefore very important to maintain continual screening of cultured cells for infection. Mycoplasma may be detected by fluorescent Hoechst 33258 staining [22], culture methods [23] or PCR-based screens [24–26]. Centralized cell services often offer mycoplasma screening. Infected cultures should be discarded.

14.5.3 Fusogens

Cells grown in culture fuse spontaneously, but at a very low frequency. Treatment of cells with inactivated Sendai and some other viruses increases fusion efficiency and was once widely used to promote cell fusion *in vitro*. However, Sendai virus is ineffective with some kinds of cells, virus production is laborious and subject to batch variation, and the virus may not be completely inactivated. The chemical fusogen polyethylene glycol (PEG) has

replaced Sendai virus as the fusogen of choice [27, 28], in spite of its potential toxicity to cells in culture. PEG is a polymer available in a range of molecular weights; cell fusion experiments are commonly performed using a single molecular weight PEG within the range of 1000–6000 Da. The size of the polymer is related to both fusion efficiency and toxicity [29]. Lower molecular weights in this range are more fusogenic but also more toxic. The decreased toxicity of higher molecular weight PEG is offset by its high viscosity which makes it difficult to wash off the cells quickly, increasing exposure and toxic effects.

The optimal PEG concentration for fusion of a particular cell type is not predictable, but usually

lies within the range 45–55% (w/w) [29]. Because of the toxicity, most fusions are performed with 45–50% PEG. The PEG concentration with the highest fusion efficiency for any particular combination of cells should be determined empirically. A good starting point is PEG 1500 at 50%, fusion-tested solutions of which are commercially available (Boehringer Mannheim Biochemica; Sigma Chemical Company). Fusion frequencies for interspecific whole cell and radiation hybrid fusions should range from one hybrid in 10^4 to one in 10^6 cells fused. The yield from microcell fusion frequencies will be lower, ranging from one hybrid in 10^5 to one in 10^7 cells fused.

.....

Troubleshooting

PEG toxicity

Observe cells carefully for the toxic effects of PEG. If many dead cells are seen following PEG exposure, steps should be taken to reduce toxicity.

- *Initially try reducing the PEG concentration.*
- *If fusion efficiency is unacceptably low, test a higher molecular weight PEG (PEG 4500, then PEG 6000), then try a different batch or manufacturer.*

Koch-Lite PEG (NBS Biologicals) has a reduced cytotoxicity compared with that from other sources [30] and does not seem to have the batch variation that is found with some other brands. Incubation of cells during and postfusion in calcium-free medium has been reported to decrease the cytotoxicity of PEG [30].

.....

14.6 Source of cells

Cells from a wide variety of sources and cell types can be fused to generate somatic cell hybrids. Hybrids have been made between cells from vertebrate species as diverse as chick [31] and mice [32], and interphyletic hybrids have been constructed [33]. In general, intraspecific crosses generate hybrids at higher frequencies than interspecific crosses, while phenotypically similar cells yield hybrids with greater efficiencies than fusion of dissimilar cell types. For this reason, adherent cell–

adherent cell fusions are preferable to the lower efficiency suspension cell–adherent cell fusions.

Hybrid morphology is intermediate to that of the parents, and will vary with the proportion of each parental genome present, a characteristic that is useful in distinguishing hybrid clones from unfused parental cells. Since the recipient cell typically supplies a full genome background providing most cellular functions, the morphology of the recipient cell is usually predominant. Therefore, suspension cells used as donors in a fusion with adherent recipient cells will most often yield adherent

hybrids. Cells containing a small proportion of donor DNA will tend to resemble the parental recipient cell.

Different criteria are used in selecting the parental recipient and the parental donor cells. If a phenotype is being studied, the recipient cells used are restricted to those displaying the phenotype, and donor cells restricted to those that modulate the phenotype. For mapping cloned DNA sequences, the source of donor cells is one containing the genome, or portion of genome, to be studied. The recipient cell is usually a hardy cell line with a metabolic deficiency that can be complemented by the donor. Mouse L-cell derivatives, such as the cell lines L-M TK⁻ (TK⁻) [34] and A9 (HPRT⁻, APRT⁻) [35] are often used, in spite of being hyperdiploid and having a complex karyotype. Two other widely used recipient cell lines are the Chinese hamster derivatives Wg3H (HPRT⁻) and A23 (TK⁻) [36]. The cells grow very well and the low chromosome number in the hamster lines ($2n=22$) facilitates karyotypic analysis. Generation of somatic cell hybrids requires that the recipient cell line divides in culture and usually an immortalized cell line is used in order to establish permanent hybrid lines. Primary cells with a limited lifespan have been used as recipients in fusions, for example in studies of cellular senescence.

Established cell lines may be used as donor cells, although such cells have often undergone chromosomal rearrangements, and not many diploid or pseudodiploid cell lines are available. Primary cells are less likely to have genetic changes associated with cells kept in prolonged culture and are commonly used as donor cells in fusions. Primary skin fibroblasts are easily obtained, generally grow well (during their limited lifespan) and fuse well. Other tissues may serve as a source for donor cells, although obtaining a pure population in quantities sufficient for fusion may be problematic. MMCT requires donor cells that divide in culture, even if only once, to form the micronuclei necessary for the fusion.

Several large cell repositories collectively offer thousands of cell cultures available at reasonable cost. The most extensive are the European Collection of Animal Cell Cultures (ECACC, Salisbury, UK), the NIGMS Human Genetic Mutant Cell Repository, and the American Type Culture Collection (ATCC, Rockville, MD, USA) (see Appendix V for addresses). These collections are a source of a wide variety of cells useful for hybrid construction including karyotypically simple human-rodent hybrids, primary cells from patients, and many phenotypically diverse cell lines. Cell lines com-

monly used as recipients in fusions are available as well.

14.7 Chromosome segregation

A general property of somatic cell hybrids is the loss (segregation) of chromosomes derived from one or both of the parental genomes. Chromosome loss initially occurs fairly rapidly, after which the chromosome number stabilizes and chromosome segregation is slow [37].

Intraspecific crosses tend to retain most chromosomes from both parents and chromosomes from either parent can be lost. Chromosome loss in interspecific hybrids can be extensive. In most instances, chromosomes of one species are preferentially eliminated and the full complement of chromosomes from the other species is retained. The direction of loss can be influenced by several factors, the most important of which are (i) the particular species combination, and (ii) whether one parent is a primary cell isolate. Species is the important determinant of segregation in crosses of permanently established cell lines. Ordinarily, in fusions of established cells, human-rodent hybrids segregate human chromosomes, rat-mouse hybrids segregate rat chromosomes, and mouse-hamster hybrids segregate mouse [38], although segregation can be bidirectional in rodent interspecific cell hybrids.

The culture status of the cells fused most often overrides the species importance. Fusion of primary diploid cells with an established cell line nearly always results in the loss of chromosomes from the primary cell line [39]. This effect may be a result of the propensity for the chromosomes from the faster-growing parent to be retained. Segregation can be directed to a certain extent by damaging the chromosomes of the parental cells prior to fusion with X- or gamma-rays, treatment with 5-bromodeoxyuridine (BrdU) or by direct selection against a marker on a specific chromosome [38]. The particular chromosomes retained stably in a hybrid are unpredictable, so if retention of a specific chromosome is desired, it is best to ensure its presence by direct selection on an endogenous or integrated marker present on the chromosome (Table 14.2).

Overall, the best approach to controlling the loss of a particular species chromosome is to fuse diploid primary cells with an established rodent cell line. In most instances the chromosomes from the primary cell parent will be segregated from the hybrids.

14.8 Selection

Biochemical selection methods are employed in the

Table 14.2 Endogenous biochemical selection genes.

Human chromosome	Selected marker	Forward selection	Reverse selection	Rodent recipient	Reference
<i>Chromosome 1</i>					
p35-p36.2*	Deoxycytidine deaminase (CDA)	Hypoxanthine + aminopterin + 5-methyldeoxycytidine (HAM)	5-bromo-deoxycytidine	3T6-BCE (CDA-DCK-)	[45,147]
p34.1-p34.3*	CTP synthetase (CTPS)	Minus cytidine		CHO CR-2	[148,149]
p22.2-qter*	Succinate dehydrogenase (SDH)	DME-GAL (glucose replaced by galactose)		CCL16-B9 (Chinese hamster)	[150]
p13*	Adenosine monophosphate deaminase (AMPD1)				[151]
q*	Adenylosuccinate synthetase (ADSS)	Purine-free		CHO-K1, Ade-H	[152,153]
<i>Chromosome 2</i>					
2*	AICAR formyltransferase	Purine-free		CHO-K1, Ade-F	[154]
<i>Chromosome 3</i>					
q13	Uridine monophosphate synthetase (UMPS) (orotate phosphoriboyl transferase and orotidine- 5'-decarboxylase)	Uridine-free		CHO-K1, Urd-C	[155]
<i>Chromosome 4</i>					
p16.3-q21*	Phosphoribosyl pyrophosphate amidotransferase (PPAT)	Hypoxanthine-free		CHO-K1, Ade-A	[156,157]
q11-qter*	Phosphoribosyl aminoimidazole carboxylase phosphoribosylamino-ribosylaminoimidazole succinocarboxamide synthetase (PAICS)	Hypoxanthine-free		CHO-K1, Ade-D	[158]
<i>Chromosome 5</i>					
cen-q11	Leucyl tRNA synthetase (LARS)	39°C		UCW 56 (CHO, emtBr, leuS ^{ts} and chr ^r) CHO dhfr ⁻	[159] [160-162]
q11.2-q13.2	Dihydrofolate reductase (DHFR)	MM (minus glycine, purines, and thymidylate)			
q23	Diphtheria toxin receptor (DTS)		Diphtheria toxin	Mouse cell lines	[163,164]
q31-q33	Ribosomal protein S14 (RPS14)		Emetine	UCW 56	[165]
q35	Chromate resistance; sulphate transport (CHR)		Sodium chromate	UCW 56	[159,166]
<i>Chromosome 6</i>					

Continued on p. 330.

Table 14.2 Continued.

Human chromosome	Selected marker	Forward selection	Reverse selection	Rodent recipient	Reference
Chromosome 7 q21-q31	Asparagine synthetase (ASNS)	Asparagine-free		N3(CHO, ASNS-)	[167,168]
Chromosome 8 q21.1 qter*	Glycine B complementing (GLYB)	Glycine-free		CHO Gly-B	[169]
Chromosome 9 q12-pter	Methylthiosine phosphorylase (MTAP)	MM + azaserine + methylthio adenosine(MTA) (MAM)		Various mouse and human tumour MTAP ⁻ cell lines	[170]
cen-q34	Folyolyglutamate synthase (FPGS)	Minus glycine, adenine and thymidine		CHO-K1, GAT ⁻	[171,172]
Chromosome 10 cen-q24	Adenosine kinase (ADK)	Adenosine or 2-fluoroadenosine (FAR) + alanosine + uridine	(i) Toyocamycin (ii) Tubercidin (iii) 6-Methylthio-purine riboside (iv) 2-Fluoro-adenosine (v) Adenosine	FR5 (mouse DF8 APRT ⁻ , ADK ⁻)	[173]
10*	Glutamate γ-semialdehyde synthetase (GSAS)	Proline-free		CH, Pro ⁻	[174]
Chromosome 11					
Chromosome 12 q12-q14	Serine hydroxyl-methyltransferase (SHMT)	Glycine-free		CHO-K1, Gly-A	[59,175]
Chromosome 13					
Chromosome 14 14*	Phosphoribosyl-formylglycine-amide amido-transferase (PFGS)	Purine-free		CHO-K1, Ade-B	[176]
	Methylenetetrahydrofolate dehydrogenase-methenyltetrahydrofolate cyclohydrolase-formyltetrahydrofolate synthetase (MTHFD)	Purine-free		CHO-K1, Ade-E	[177,178]
Chromosome 15					
Chromosome 16 q24.2-qter	Adenine phosphoribosyltransferase (APRT)	(i) Alanosine + adenine (AA) (ii) Adenine + aminopterin + thymidine (AAT) (iii) Adenine + azaserine	(i) 2-Fluoroadenine (FA) (ii) 8-Aza-adenine (iii) 2,6-Diamino-purine (DAP) (iv) 6-Mercaptopurine	Various e.g. A9; 585MEL	[44,179,180]

Continued.

Table 14.2 Continued.

Human chromosome	Selected marker	Forward selection	Reverse selection	Rodent recipient	Reference
Chromosome 17 q23.2-q25.3	Thymidine kinase 1 (TK1)	Hypoxanthine + aminopterin (or methotrexate) + thymidine (HAT)	(i) 5-Bromo-deoxy-uridine (BrdU) (ii) 5-Fluoro-deoxy-uridine (iii) 5-Iodo-deoxy-uridine (iv) Trifluoro-thymidine	Various e.g. A23 (Chinese hamster)	[36,40–42]
Chromosome 18 q11.31-p11.21	Thymidylate synthetase (TYMS)	Thymidine-free	Folinic acid + aminopterin + thymidine + cyanocobalamin (FAT)	V79 TYMS ⁻ (Chinese hamster)	[181–183]
q21.2–q21.3*	asparaginyl-tRNA synthetase (NARS)	39°C and asparagine-free		Asn-5 (CHO, NARS ⁻)	[184]
Chromosome 19					
Chromosome 20 q12-q13.11	Adenosine deaminase (ADA)	Deoxyadenosine		ADA	[185]
Chromosome 21 q22.1	Phosphoribosyl-glycinamide formyltransferase, cinamide synthetase, phosphoribosyl-glyphosphoribosyl-aminoimidazole synthetase (GART)	Purine-free		CHO-K1, Ade-C CHO-K1, Ade-G	[186–188]
Chromosome 22 q13.1	Adenylosuccinate lyase (ADSL)	Adenine-free		CHO-K1, Ade-I	[189]
Y chromosome					
X chromosome p11.23	Ubiquitin-activating enzyme E1 (UBE1/A1S9T)	39°C		tsA1S9 (a UBE1-mouse cell line)	[190]
q26	Hypoxanthine phosphoribosyl-transferase (HPRT)	Hypoxanthine + aminopterin (or methotrexate) + thymidine (HAT)	(i) 6 - Thioguanine (6-TG) (ii) 8-Azaguanine (8-AG)	Various e.g. Wg3h (Chinese hamster); A9 (mouse fibrosarcoma)	[36, 40, 42, 191]
Map location unknown	Deoxycytidine kinase (DCK)	HAT + deoxy-cytidine	Cytosine arabinoside (araC)	CHO, araC ⁻ 3T6-BCE (CDA DCK)	[45,192,193]

* Provisional map location. MM, minimal essential medium.

generation of somatic cell hybrids for two purposes: first, to kill one or both of the unfused parental cells, enabling the efficient recovery of hybrid clones; second, to enrich for or against cells containing a chromosome (or portion of the chromosome) that carries the selectable marker.

14.8.1 Endogenous selection genes

In the first hybrid isolation experiments, mutant cells were used as recipients. The fusion was monitored by the transfer of genes that complemented the deficiencies and could be selected for in media toxic to the host cells, allowing only hybrid cells to grow. The variant cell lines most widely used in the isolation of somatic cell hybrids are those resistant to the drugs 8-azaguanine (8-AG)/6-thioguanine (6-TG) or 5-BrdU. The widespread use of these mutants has been associated with their role in the HAT biochemical selection system (see Section 14.8.1.1).

Endogenous markers can also be used to select for retention of a donor chromosome. In principle, any human chromosome can be selected for if an appropriate mutant recipient cell line is available for that chromosome.

14.8.1.1 HAT selection

In the early 1960s Szybalski and colleagues [40,41] showed that it is possible to obtain mutant cells

defective in specific enzymes by subjecting a normal cell population to selection with drugs. This observation formed the basis of a general method for the isolation of hybrid cells. Drug-resistant cells will arise spontaneously in cell cultures, although the frequency can be increased by the use of X-irradiation or the chemical mutagens 8-AG or 6-TG to kill normal cells. When these drugs are metabolized they interfere with normal nucleotide and nucleic acid synthesis (Fig. 14.1). This effect is mediated by the enzyme hypoxanthine phosphoribosyltransferase (HPRT), which converts the drugs to ‘abnormal’ nucleotides. Mutant cells lacking HPRT are resistant to the drugs and therefore survive treatment. BrdU-resistant cells can be obtained by a similar procedure. In normal cells this drug will first be phosphorylated by thymidine kinase (TK) and then incorporated into DNA. This normally results in cell death. Mutant cells defective in TK fail to phosphorylate and incorporate BrdU into the DNA and are therefore drug resistant.

These genetic defects are of little importance during growth in normal tissue culture media, as these enzymes are only involved in salvage pathways for nucleotide synthesis. However, such HPRT⁻ or TK⁻ cells cannot grow in HAT medium, which contains hypoxanthine, aminopterin (or methotrexate) and thymidine. This is because aminopterin, a folic acid analogue, blocks *de novo* synthesis of purines and pyrimidines and in the absence of HPRT or TK

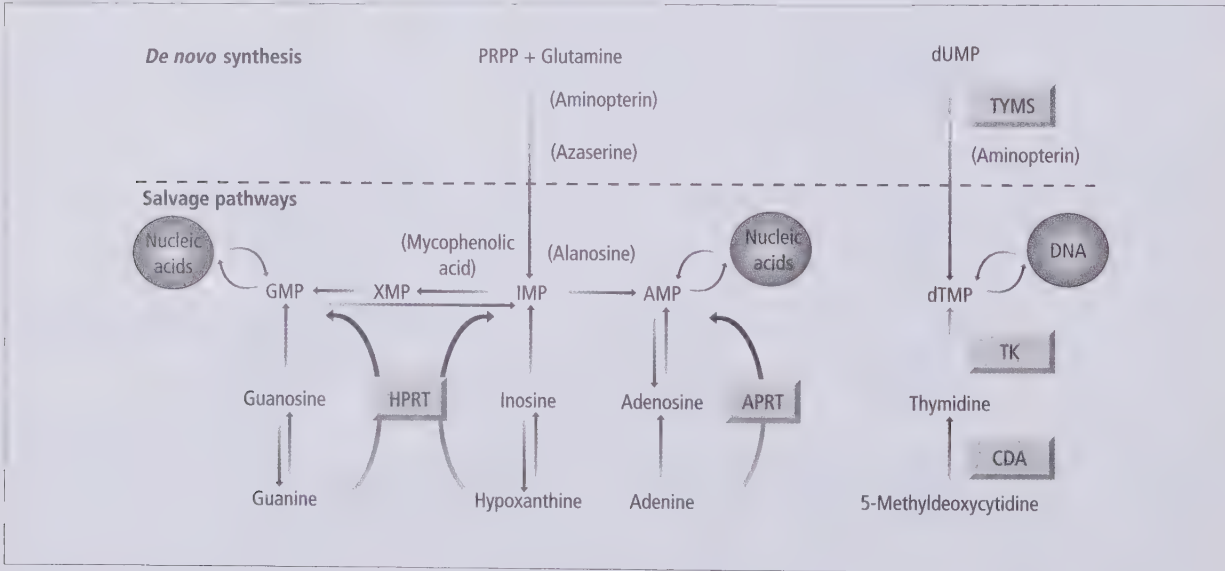


Fig. 14.1 Salvage pathways to obtain purine nucleotides and thymidylate. Blocking of synthesis is indicated by parallel lines. AMP, adenosine 5'-monophosphate; APRT, adenine phosphoribosyltransferase; CDA, deoxycytidine deaminase; dTMP, deoxythymidine 5'-monophosphate; dUMP, deoxyuridine 5'-monophosphate; GMP, guanine

5'-monophosphate; HPRT, hypoxanthine phosphoribosyltransferase; IMP, inosine 5'-monophosphate; PRPP, 5'-phosphoribosyl-1-pyrophosphate; TK, thymidine kinase; TYMS, thymidylate synthetase; XMP, xanthosine 5'-monophosphate.

activity the cells cannot utilize the exogenous compounds via salvage pathways.

In 1964, Littlefield [42] showed that when 8-AG-resistant (HPRT⁻/TK⁺) and BrdU-resistant (HPRT⁺/TK⁻) mouse fibroblasts are mixed in HAT medium, cells of neither parental cell line survive, but hybrids between the two lines are able to grow. Thus, when combined in one cell the two parental genomes complement each other. Many HPRT⁻ and TK⁻ cell lines have been derived from human, mouse, rat, Syrian hamster and Chinese hamster cells and these lines have been used extensively in a variety of hybridization experiments.

A common variant of the HAT theme is 'half-selection' where one cell type does not grow in culture (or grows more slowly) and the other parent is HPRT⁻ or TK⁻ deficient. Therefore one can select against a rodent parent using non-permissive medium and against a human diploid cell by virtue of its inherently poor growth characteristics.

Such selections were initially used to generate hybrid panels in which the various lines retained one or a few human chromosomes, and to recover monochromosomal hybrids for the specific chromosomes bearing these particular loci, or occasional naturally occurring translocation chromosomes (where the marker has been transferred onto another chromosome) [43]. More recently, HAT selection has been used extensively to rescue the fusion products generated in IFGT experiments (see Section 14.12). Upon removal of HAT selection, cells need to be cultured in transition medium (hypoxanthine-thymidine; HT) for a few days to allow all residual aminopterin to be cleared.

HAT and HT supplements are available commercially (Gibco-BRL) or can be prepared as in Protocol 68.

14.8.2 Biochemical selections for other endogenous markers

In addition to the HAT selection system there are now other selection methods based on the fusion of drug-resistant cells and selection in special media (Table 14.2). Kusano, Long and Green [44] showed that cells deficient in adenine phosphoribosyl-transferase (APRT) could be obtained after selection in fluoroadenine or 6-aminopurine. Hybrids between APRT⁻ mouse cells and APRT⁺ human fibroblasts could be obtained by selection in medium containing adenine plus alanosine (AA) (alanosine is an antibiotic which prevents the formation of AMP from inosine monophosphate (IMP), and APRT⁻ cells cannot survive in adenine + alanosine medium) or in AAT medium containing adenine,

thymidine and aminopterin. Chan *et al.* [45] showed that cell lines deficient in deoxycytidine kinase (DCK⁻) and deoxycytidine deaminase (CDA⁻) are unable to grow in HAM media (hypoxanthine, aminopterin and 5-methyldeoxycytidine).

It is also possible to select for hybrids between HPRT⁻ and APRT⁻ mutant cells using a selective medium called GAMA [46]. GAMA contains azaserine (to block the endogenous synthesis of purines; mycophenolic acid (to block the conversion of AMP to GMP); and guanine and adenine as the sole purine sources. Cells in GAMA medium rely solely on the supplemented adenine and guanine for their adenine and guanine nucleotides and therefore require both HPRT and APRT enzyme activities.

These genes are located on human chromosomes X (HPRT), 17 (TK) and 16 (APRT). Similarly, active selection can be applied for mouse chromosomes X (HPRT), 11 (TK) and 8 (APRT).

An effective way to remove human donor cells in cell hybridization experiments is by adding ouabain to the medium. This is based on the observation that rodent cells are at least 10 000-fold more resistant to ouabain toxicity than are human cells in culture [47]. Since the human-rodent hybrid will show a level of ouabain resistance intermediate to the two parental cell types, the ouabain concentration should be just sufficient to kill the human donor cells. Ouabain is thought to act by inhibiting the Na⁺/K⁺ ATPase.

14.8.3 Other selection schemes for endogenous markers

Some chromosomes express genes for which other forms of positive selection can be applied. Morphologically transformed phenotypes have been used as a positive selection factor and cell hybrids specifically carrying human chromosome 11 can be selected for by the *ras* oncogene [48] and human chromosome 7 by the *met* oncogene [49] using appropriate recipient cells. Similarly, selection can be for genes conferring tumour suppressor activity [6,7,50,51], cellular senescence [5,52] or complementing DNA repair defects [9,53,54].

The identification of cell-surface antigens has also been instrumental in mapping chromosomes. Monoclonal antibodies can be used either to isolate (panning, FACS) or to select against (complement-mediated cell lysis) antigen-positive cells [55–57] (see Chapter 18).

14.8.4 Hybrids made from auxotrophic mutants

Conditional lethal mutants other than those based on drug resistance can also be used for hybrid

selection. Auxotrophic mutants have been isolated by screening cells for inability to grow on medium lacking some metabolite, contrasted with ability to grow on complete medium (e.g. the BrdU-killing method). Kao and Puck [58–60] selected several classes of auxotrophs from Chinese hamster ovary (CHO) cells. These included glycine-requiring mutants, adenine variants, ino^- cells requiring inosine, and some with double requirements for adenine and thymine (AT^-) or triple demands for glycine, adenine and thymidine (GAT^-). Many of the complementing human enzymes have now been mapped to specific chromosomes (see Table 14.2). As a result, for many human chromosomes, a suitable mutant rodent recipient cell line has already been generated that can be used for the isolation of that chromosome onto a rodent background (Table 14.2). This approach is of particular value where the requirement is to analyse an abnormal chromosome associated with a particular syndrome.

14.9 Insertion of dominant selectable markers into the mammalian genome

A lack of selectable markers can be obviated by the introduction of exogenous genes for active selection. Mutant cell lines can be used for the introduction of non-dominant markers, such as *HPRT*, *APRT*, *TK* and *DHFR*, at heterologous chromosomal sites. Alternatively, foreign genes introducing new dominant characters into normal recipient cells can be used. A generally applicable method for isolating hybrids is to ‘tag’ the chromosome of interest with an exogenous dominant selectable marker. Although in theory the marker can be introduced using homologous recombination [61,62], more frequently the route is via random integration. Several prokaryotic antibiotic-resistance genes have been isolated and vectors for their efficient expression in mammalian cells have been constructed. The approach of introducing a dominant selectable marker, such as *neo*, *hisD* or *gpt*, into diploid human fibroblasts, and transferring the marked chromosomes to rodent recipients by microcell fusion, has been applied by a number of groups and several such tagged monochromosomal human–rodent hybrids have been described in the literature (Table 14.3). The most widely used selection schemes are described in Section 14.9.2. Selection is usually based on resistance to a substance that is toxic to normal cells.

14.9.1 Dominant selectable markers

The most useful cell selection systems are those that

do not require mutant recipient cells, but can be used with any type of recipient cell. Several such systems have now been developed.

For any selection strategy, the optimal conditions for any particular combination of cell line and marker must first be established. A titration curve of the selective agent must be produced in order to establish the lowest effective killing concentration, and to determine whether spontaneous resistance arises in the recipient cells, and if so at what frequency. Different cell types can exhibit substantially different sensitivities to the various selective drugs.

14.9.1.1 Transfer of dominant selection genes into mammalian cells

Such markers can be introduced through a variety of gene transfer techniques, such as the coprecipitation of the DNA with calcium phosphate or alternative cationic agents, electroporation, lipofection and microinjection. Integration of foreign DNA is random and may involve either one or several copies at single or multiple integration sites, depending on the recipient cell type and on the transfection method used. Calcium phosphate precipitation often results in complex integration events, whereas following electroporation many transfectants display one single-copy integration. Integration of genes into chromosomes can also have a destabilizing effect. Using retroviral vectors rather than plasmids, it is possible to stably transform up to 100% of a cell population, and retroviral infection normally results in the integration of a single copy of the viral genome in each cell [63].

14.9.1.2 Eukaryotic expression vectors

Another crucial factor determining transfection frequencies will be the choice of promoter driving the selectable marker. The activity of the promoter/enhancer varies with the host cell type, probably due to dependence on cell-specific transregulatory factors. Depending on the type of experiment being undertaken, such differences may be crucial in determining success or failure.

The prototype eukaryotic expression vector pSV₂ is based on simian virus 40 (SV40) control sequences for gene expression. It consists of the SV40 replication origin, containing the early gene promoter, enhancer and transcription start point; the small t antigen intron, and the small t antigen polyadenylation site [64]. As the SV40 promoter functions in a large number of cell types, this vector has been extensively used to introduce dominant selection genes into mammalian cells. Widely used alternatives to the SV40 promoter/enhancer system are the long terminal repeats from Rous sarcoma

Table 14.3 MMCT-generated human monochromosomal somatic cell hybrids described in the literature.

Selectable marker	Human chromosome	Donor	Recipient	Reference
<i>neo</i>				
(i)	2/3/5/15	CSC301 (primary diploid fetal lung-derived fibroblasts)	A9 (mouse)	[194]
(ii)	4/11/14/20/21	Primary diploid human fibroblasts	NIH 3T6 (mouse)	[195,196]
(iii)	17	Primary diploid human fibroblasts	La-t-(APRT ⁻ , (TK-mouse cell line)	[102] [197]
(iv)	5/11der/20		FM7 (Indian muntjac)	
(v)	1/2/5/6/7/8/ 9/10/11/12/15/ 17/18/19/20	MRC-5 or NTI-4 (primary human fibroblasts)	A9	[51,198]
<i>hisD</i>				
(i)	5/9/12/19	GM 7890 and GM 7965 (human lymphoblastoid cell lines)	CHO UV-135	[78]
<i>gpt</i>				
(i)	17	HFL121 (human fibroblasts)	1R (mouse L cells)	[106]
(ii)	2/5/6/9/12q/ 13/16/17/21		A9 and CHTG-49 (Chinese hamster)	[54,108,199]
(iii)	2/4/22	HT1080 and D98/AH-2 (human HPRT ⁻ cell lines)	A9	[200]
<i>hyg</i>				
(i)	11	HR9, M (h11) (mouse-human hybrids)	DT40 (chicken)	[79]
<i>HyTK</i>				
(i)	all 22 autosomes/X	1BR.2 (human diploid fibroblasts) or pre-existing hybrids	A9	[201]

This table does not aim to be comprehensive. Many more MMCT-monochromosomal hybrids have been generated in the course of studies on genes for which functional assays are available, e.g. DNA repair defects, cellular senescence and tumour suppressor activity. Moreover, some MMCT-generated monochromosomal hybrids have also been described for mouse and other species [43,202].

virus (RSV) and the cauliflower mosaic virus (CMV) promoter. A range of expression vectors constructed using promoters, enhancers and intron sequences of various origins is now available [65,66].

14.9.2 Selection schemes

14.9.2.1 Positive selection

In the following schemes (1–9) the name of the drug used for selection is given first, followed by the enzyme or other protein whose presence is being selected for, with the name of the corresponding gene in brackets.

1 G418 sulphate (G418)/neomycin phosphotransferase II (*neo*) The neomycin phosphotransferase II (NPTII) gene (*neo*) is derived from the *Escherichia coli* transposon Tn5 and was first described as conferring resistance to the aminoglycoside antibiotic G418 sulphate in yeast [67]. G418 is similar to neomycin and kanamycin and acts by inhibiting protein

synthesis, specifically by interfering with the function of 80S ribosomes. Spontaneous resistance is very low. The NPTII gene protects from G418 toxicity by phosphorylating the drug, thereby inactivating it. Cells need therefore to be actively dividing for G418 to exert its effect, with the most rapidly growing cells being killed in the shortest interval. With fibroblasts, detachment (indicating cell death) should take place within 5–10 days of treatment.

For efficient expression, the *neo^r* gene has been placed under the control of regulatory elements from different viral genes such as the *tk* gene [68,69] from the herpes simplex virus (pMC1neo) (Stratagene), the SV40 early promoter (pSV2neo) [70] or the RSV long-terminal repeat (LTR). The *neo^r* gene has also been used in the retroviral vector pZIP-NeoSV(X)1 [71]. pSV2neo and pMC1neo have been shown to confer G418 resistance on a variety of cell types, including mouse (NIH3T3, ES cell lines, LMTK⁻ and 3T6), monkey (Vero, OMK and TC7) human (HeLa, HT1080), hamster (Wg3H), and

chicken (DT40). We have found that this selection system has disadvantages when working with some rodent cell lines (NIH3T3 and Chinese hamster). Although these lines are fast growing, with doubling times of around 16 h, G418-induced killing is slow (10–14 days) and transformation frequencies using pSV2 neo or pMC1neo are low (1 in 10⁴ or 10⁵ recipients).

Preparation G418 sulphate is available commercially as Geneticin (Gibco-BRL). It should be noted that details of G418 concentrations usually refer to the concentration of active drug, rather than to the G418 concentration (the active drug concentration is about 40–50% the concentration of G418). The killing effect of G418 may vary slightly between batches and if details of the active drug concentration are not available fresh titration curves will need to be performed. G418 is supplied as a powder and should be dissolved at a concentration of 100 mg ml⁻¹ in a highly buffered solution (e.g. 0.1 M Hepes, pH 7.3) so that addition of the drug does not alter the pH of the medium. Its stability in solution is good. Store at 4 °C. Working concentrations of the active drug generally range from 100 µg ml⁻¹ to 1 mg ml⁻¹. A disadvantage of this selection scheme is that relatively large amounts of G418 are needed and it is expensive.

2 Mycophenolic acid/xanthine phosphoribosyltransferase (*gpt*) The *gpt* gene from *E. coli* encodes the enzyme xanthine phosphoribosyltransferase (XPRT) which, when expressed, confers resistance to mycophenolic acid on animal cells. This scheme exploits the fact that mammalian cells do not possess an equivalent enzyme that can utilize xanthine. The selection is best carried out in medium containing aminopterin (to block *de novo* purine synthesis), mycophenolic acid (an inhibitor of IMP dehydrogenase), and supplemented with adenine, thymidine and xanthine [72,73].



A drawback of this selection scheme is that it suffers from leakiness, with a significant proportion of some wild-type cells surviving selection. Moreover, selection with mycophenolic acid requires dialysed serum, as well as medium formulated without guanine.

In addition to being useful as a dominant positive selectable marker, *gpt* can also rescue HPRT-deficient cells from HAT sensitivity. Moreover, back

selection is possible by selection in 8-AG, 6-TG or 6-thioxanthine (the latter being specific for *gpt*).

Preparation The mycophenolic acid (Sigma) is made up as a 100× stock (2.5 mg ml⁻¹) in 0.1 N NaOH, neutralized with 0.1 N HCl. The xanthine and hypoxanthine are prepared as a 100× stock as follows: 25 mg ml⁻¹ xanthine (28.6 mg ml⁻¹ of the sodium salt) plus 1.5 mg ml⁻¹ hypoxanthine, in 0.3 M NaOH.

To make *gpt*-selective medium add the following to medium, without guanine: dialysed FCS; 250 µg ml⁻¹ xanthine; 15 µg ml⁻¹ hypoxanthine or 25 µg ml⁻¹ adenine; 10 µg ml⁻¹ thymidine; 2 µg ml⁻¹ aminopterin; 25 µg ml⁻¹ mycophenolic acid and 150 µg ml⁻¹ L-glutamine.

3 Hygromycin B/hygromycin B kinase (*hph*) The prokaryotic drug-resistance gene *hph* encoding hygromycin B kinase is widely used as a dominant selectable marker [74]. Hygromycin B is an aminocyclitol antibiotic that acts by inhibiting protein synthesis, interfering with ribosomal translocation and causing mistranslation. The *hph* gene confers hygromycin resistance on a variety of human, mouse (e.g. NIH-3T3), hamster (e.g. Wg3H) and chicken (DT40) cell lines. In its absence, most cell types are rapidly killed by hygromycin, with the majority of cells dead and/or detached within 48 h of exposure. Spontaneously resistant colonies are very rare.

Preparation Hygromycin B is available in liquid form (Calbiochem) and is stable at 4 °C. Working concentrations are in the range 100 µg ml⁻¹ to 1 mg ml⁻¹. We have found that Wg3H Chinese hamster fibroblasts do not recover well from freezing after growth in hygromycin B-containing media. It is not known whether this is also the case for any other cell types.

Because of the different specificities of the two antibiotics G418 and hygromycin B, these selection schemes can be used simultaneously and independently in lines that express both *neo* and *hph*.

4 Puromycin/puromycin N-acetyl transferase (*pac*) Puromycin is another antibiotic that blocks protein synthesis by 80S ribosomes. The *pac* gene from *Streptomyces alboniger* is useful in selecting genetically transformed mammalian cells (monkey, Vero; mouse, L and NIH-3T3; hamster, BHK21; and human, HeLa) [75,76]. We have found that with Chinese hamster fibroblasts, this drug is effective over a very restricted concentration range (below this concentration the frequency of spontaneous puromycin-resistant colonies is significant, whereas at slightly higher concentrations no transformants

were recovered following electroporation of a pSV2-pac construct).

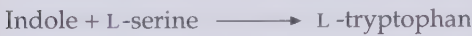
Preparation Dissolve puromycin powder (Sigma) as an aqueous stock at 10 mg ml⁻¹. Adjust pH to 9–10 using NaOH. Working concentrations are in the range 1–100 µg ml⁻¹.

5 Histidinol/histidinol dehydrogenase (*hisD*) The histidinol dehydrogenase gene (*hisD*) of *Salmonella typhimurium* enables direct selection of transformed cells in histidinol [77]. Histidinol is toxic to mammalian cells because it interferes with protein synthesis by competing with histidine for histidyl-tRNA synthetase. Histidinol dehydrogenase converts L-histidinol to L-histidine, thereby protecting the cells.

HisD⁺ transformants can also be selected under histidine-minus selection conditions, but this requires specially prepared medium. *hisD* has been reported to function as a selectable marker in mouse NIH-3T3 cells, monkey CV-1 cells, murine embryonal stem cell lines, human HeLa cells, Chinese hamster Wg3H 61, human lymphoblastoid and chicken (DT40) cell lines [78,79].

Preparation L-histidinol is supplied as a powder (Sigma) and should be prepared as a 100 or 200 mM aqueous stock stored at -20 °C until required. This selective agent is effective against hamster and mouse cell lines at a final concentration of 5 mM (in complete medium, i.e. containing histidine). Cell death occurs over a period of about 4–7 days. Higher concentrations are required for effective killing of established human cell lines, such as HT1080. It is worth noting that although histidinol appears less expensive, at the concentrations usually required for effective killing, the final cost works out similar to that for G418 sulphate.

6 Tryptophan/tryptophan synthase (β -subunit) (*trpB*) Expression of the *E. coli trpB* gene enables mammalian cells to survive in the presence of indole when tryptophan is excluded from the culture medium. The *trpB* gene encodes the β -subunit of tryptophan synthase, which catalyses the reaction:



trpB has been used as a selectable marker in mouse NIH-3T3 cells, monkey CV-1 cells, and human HeLa cells. A disadvantage of this selective agent is that specially prepared tryptophan-deficient medium is required [77].

7 Phleomycin/bleomycin- and phleomycin-binding protein (*ble*) Bleomycin and phleomycin are glycopeptide antibiotics that kill mammalian cells at concentrations of 5–10 µg ml⁻¹ by causing site-specific breaks in DNA [80].

The *ble* gene of *Streptoalloteichus hindustanus* encodes a binding protein with high affinity for both antibiotics. When bound in a complex with this protein, neither pleomycin nor bleomycin can be activated by ferrous ions and oxygen to react with DNA. The *ble* gene has been shown to confer resistance to the drug in CHO cells and in NIH-3T3 cells [76,81]. This selection system is now available commercially under the trade name zeocin (Invitrogen).

8 Albizzin/asparagine synthetase (*asnA*) The bacterial asparagine synthetase (AS) gene, which catalyses the formation of asparagine from aspartic acid, has been used to derive albizzin-resistant transfectants of wild-type AS⁺ cell lines of rat, Chinese hamster, mouse and human origin [82] (as well as complementing AS⁻ cell mutants). Selection is in medium lacking asparagine. Like G418, albizzin is relatively expensive. The efficiency of pSV₂AS transfection in cells other than CHO has been reported to be relatively low [83] and this selection scheme has not been widely used by others. NB. Like DHFR (see 9 below) this marker can also be amplified to high copy number.

Preparation Dissolve albizzin powder (Sigma) as a 200 mM aqueous stock. Working concentrations are in the range of 2–10 mM.

9 Methotrexate/dihydrofolate reductase (*DHFR*) The enzyme dihydrofolate reductase (DHFR) catalyses the reduction of dihydrofolate to tetrahydrofolate, which is required for single-carbon transfers in the synthesis of glycine, purines and thymidylate. Introduction of a *DHFR* gene into DHFR⁻ CHO cells enables isolation of DHFR⁺ clones. The drug methotrexate (MTX), a folate analogue, kills cells by binding to the catalytic site of DHFR. Some cells become resistant to the drug by amplifying the *DHFR* gene so that more enzyme is produced per cell, while others become resistant because the DHFR enzyme they express has mutated so that it has decreased affinity for the drug. Mutant *DHFR* genes provide markers that can be selected by MTX resistance after the DNA is transfected into a wide variety of wild-type cells [83,84]. Moreover, if the transfectants are exposed to a gradual escalation in exposure to MTX, amplification of *DHFR* gene copy number is selected for [85]. However, it has been found that although the mutant DHFR exhibits reduced affinity for MTX, resistance is often obtainable over a very narrow range of MTX concentration, making drug resistance in this system strongly subject to the level of *DHFR* gene expression.

10 *Blasticidin/blastidicin S deaminase (bsd)* Blasticidin S deaminase and blasticidin S acetyltransferase (bsr) are enzymes which convert blasticidin S to a non-toxic derivative. Both are reported to work well in mammalian cells [86].

Other positive selection systems have been described, but most have limited application.

- The gene for adenosine deaminase (ADA) [87] when amplified provides resistance to normally toxic doses of deoxycoformycin.
- Ornithine decarboxylase (ODC) [88], when amplified provides resistance to α -methylornithine and α -difluoromethylornithine.
- CAD [89,90] is a multifunctional protein whose aspartate transcarbamylase activity can be inhibited by *N*-(phosphonacetyl)-L-aspartate (PALA). When amplified in Syrian hamster cells CAD provides resistance to normally toxic doses of PALA.
- Multidrug resistance P-glycoprotein (MDR1): The expression of *MDR1* cDNA confers resistance in a variety of mouse, human and canine cells to a variety of cytostatic drugs, among them colchicine, vinblastine and doxorubicin [91]. Cells expressing drug resistance based on MDR1 can take a long time to grow out in cultures, with non-resistant cells lingering for very extended periods.

14.9.2.2 Negative selection schemes

11 *Gancyclovir/herpes simplex virus thymidine kinase (HSV-TK)(tk)* The *tk* gene from herpes simplex virus can rescue TK-deficient mammalian cells from HAT selection [68]. Although the usefulness of this gene as a dominant selectable marker is limited to mutant cells lacking endogenous TK activity, cells expressing the viral *tk* gene can be selectively killed because it confers sensitivity to the antiviral agents gancyclovir and acyclovir [69]. The working concentration of gancyclovir (Syntex, trade name Cytovene) is in the 1×10^{-5} M range.

12 *5-Fluorocytosine/cytosine deaminase (codA)* Transfer of the bacterial gene for cytosine deaminase (CD) to mammalian cells confers lethal sensitivity to 5-fluorocytosine, as the mammalian cells are then able to metabolize cytosine to uracil, and the innocuous compound 5-fluorocytosine to the highly toxic 5-fluorouracil. Mammalian cells, unlike some bacteria and fungi, do not normally contain CD. This system has been tested in NIH-3T3 cells [92].

14.9.2.3 Positive-negative bidirectional selection systems

Markers that allow selection both for and against mammalian cells are particularly useful. They include *HPRT* either as the endogenous gene or as an

introduced minigene [93], *gpt* (see Section 14.9.2.2), *TK* (both the mammalian gene and HSV-TK) and *APRT* (see Table 14.2). In addition, two genetically engineered bidirectional selectable markers are currently available: HyTK [94] and NeoTK [95].

14.10 Whole-cell fusion

Whole-cell fusions are a simple and effective means of generating hybrids with reduced chromosomal complexity from a genome of interest in order to study cellular phenotype and map DNA sequences.

Hybridization of whole cells is the best initial experiment to assess the phenotypic changes conferred by a donor cell genome upon a recipient cell. Both intra- and interspecific hybrids can be informative, although it is much easier to identify the donor chromosomes mediating the hybrid phenotype in interspecific fusions. However, there is always the question with interspecific hybrids of whether molecules regulating a particular phenotype will function across the species barrier. This problem can be addressed by first performing intraspecific fusions to identify a phenotype conferred by one of the parents, followed by interspecific fusions to map the locus responsible. Although donor chromosomes segregate in interspecific hybrids, pools of whole-cell hybrids effectively test all donor chromosomes since the chromosome mediating a phenotypic change will be present in at least some of the hybrids.

If the phenotype is the result of gene activation, usually enough cells in the population exhibit the phenotype for it to be detected. In the case of a repression or a loss-of-function phenotype, individual hybrid cells may show the phenotype, or for changes in amounts of cellular products, a pool of hybrids may exhibit a reduction rather than a complete loss of the product. Following identification of a hybrid phenotype in pooled whole-cell hybrid populations, clonal lines can be tested for the phenotype and the donor DNA content determined by karyotyping and DNA marker analysis. A correlation between the presence or absence of donor DNA in the hybrids and the presence or absence of the hybrid phenotype is used to map the genetic locus and provides a basis for further hybrid-mapping experiments based on partial genome transfer.

Experiments aimed at mapping cloned DNA sequences utilize interspecific hybrids to take advantage of the reduction in complexity of the donor cell genome that occurs as a result of unidirectional loss of chromosomes. Whole-cell hybrids are constructed such that the genome to be mapped

segregates and clones of individual hybrid cells are analysed (Section 14.7). For phenotype mapping, karyotypic and marker analyses are used to identify retained chromosomes. A panel of characterized hybrids can then be used to map DNA sequences. Whole-cell fusion can also be used to partition abnormal chromosomes, a practice that has been of great value in cloning of human disease genes, for example to separate a rearranged chromosome from its unrearranged homologue to allow mapping of chromosomal breakpoints associated with a disease [16,96].

Whole-cell fusions can be performed on cells in suspension [97,98] or on cells attached to their growth surface [27,28,98]. Suspension fusions are usually performed when one or both of the parent cells are non-adherent. When both parents are adherent cells, they are usually mixed and plated together and fused as a monolayer. The greater fusion efficiency of monolayer fusions makes them preferable to suspension fusion. Typical whole-cell fusion efficiencies will range from 10^{-3} to 10^{-5} for intraspecific hybridizations and from 10^{-4} to 10^{-6} for interspecific fusions.

Certain controls are valuable in assessing a fusion experiment. PEG can cause clumping of cells that may persist in the presence of selection and resemble hybrid colonies. Fusion of the recipient cell line to itself without donor cells (followed by selection) provides a reference culture for comparison to the experimental fusion. This control also tests for the possibility of increased reversion of the recipient cell selectable marker induced by the fusion process. A 'mock' fusion (no PEG added) tests selection efficiency on the parental cell mixture and will assess cross-feeding.

Every fusion must be accompanied by selection controls of the donor and recipient cell lines. Each parental cell line should be plated separately into the selection used against that cell line in the fusion. These controls test the efficacy of selection and the reversion frequency for the experiment.

Protocols for carrying out whole-cell fusions of mammalian cells both as monolayers and in suspension are given in Protocol 69.

14.11 Microcell-mediated chromosome transfer

Microcell-mediated chromosome transfer (MMCT) enables the construction of hybrid cell lines with relatively simple karyotypes [99–101]. The hybrids retain few chromosomes, single chromosomes, or chromosomal fragments from a donor cell line, in a background of the full complement of recipient cell

chromosomes [102]. To construct microcell hybrids, donor cells are subjected to a prolonged mitotic block which induces micronucleation (partitioning of individual or a few chromosomes into subnuclear packets). The micronuclei are extruded from the cells by centrifugation in the presence of cytochalasin B, forming microcells (micronuclei surrounded by plasma membrane) and are then fused to intact recipient cells (Protocol 74). Hybrids are selected using a metabolic or transgenic marker located on the desired chromosome.

The selective transfer of chromosomes and the resulting low complexity of microcell hybrids makes them useful as mapping tools. In most cases, interspecific microcell hybrids are produced. For some phenotypic studies, intraspecific hybrids may be made, but the ability to easily identify transferred chromosomes by their species origin is lost. Interspecific hybrids constructed by MMCT have been used for chromosomal assignment and subchromosomal mapping of DNA sequences, as well as serving as cloning resources for DNA markers and expressed genes. In addition, many loci have been identified and mapped by a phenotype conferred upon recipient cells, including those involved in cell differentiation [4], DNA repair [8,9,103], senescence [52], tumorigenesis [6] and apoptosis [104]. Although MMCT has been commonly used to construct interspecific mammalian hybrids, particularly rodent–rodent and primate–rodent hybrids, the method has also been used to partition chicken chromosomes in human cells [105] and human chromosomes in chicken cells [79].

In theory, any chromosome can be selectively retained in a microcell hybrid, either by selection for an endogenous chromosomal marker or by random transfection of a dominant selectable marker into a donor cell population before chromosome transfer followed by identification of hybrids containing the desired chromosome [106–108]. The complexity of the donor DNA retained in microcell hybrids can be quite variable and is not predictable for any particular donor–recipient combination. Often, one or few donor chromosomes are retained, but rearrangements and fragmentation of chromosomes occurs, and preferential retention of centromeric sequences has been observed [102]. Donor cells that form large micronuclei will contribute more chromosomes to a hybrid and recipient cells are variable in their propensity to rearrange or fragment chromosomes. Because of the variability in the amount of and integrity of donor DNA retained, it is essential to characterize the genotype of microcell hybrids carefully and fully.

The donor used in MMCT can be any dividing cell

(a requirement for micronucleation), including primary cells, established cell lines, and hybrids containing a subset of the genome to be studied. The latter have been used to generate chromosome-specific deletion panels by using monochromosomal hybrids as the donors and relying upon spontaneous [102] or irradiation-induced [109] fragmentation of the donor chromosomes. Fusion efficiencies can be poor, in the range of 10^{-5} – 10^{-7} . Because of this, it is important to determine and take into account the reversion frequency of the recipient cell selectable marker. Reversion frequencies that are close to the fusion frequency will result in unacceptable amounts of non-hybrid background. Microcell fusion is technically demanding, temperamental and requires proficiency in tissue culture techniques; it is covered in Protocols 70–74. Cell requirements and a timetable for fusion are

given in Table 14.4 to facilitate organization of the various components of a microcell transfer experiment. Success depends on careful optimization and monitoring of each step, and requires patience and perseverance.

14.11.1 Micronucleation

Dividing eukaryotic cells are blocked in metaphase when exposed to colcemid, an inhibitor of microtubule polymerization. Upon prolonged treatment, the mitotic arrest is overcome and the cells re-enter interphase, forming multiple micronuclei, each containing a single or few chromosomes. For each microcell donor cell line, the optimum colcemid concentration and length of treatment must be empirically determined. The most effective concentration is usually within a narrow range for each cell

Table 14.4 Microcell fusion outline.

(a) Requirements for a 'typical' experiment	
Number of bullets per fusion	12–24
Total micronucleated donor cells	$1.2\text{--}2.4 \times 10^7$ (10^6 per bullet)
Total microcells to fuse	$0.5\text{--}2 \times 10^7$ per 25-cm ² flask of recipient cells
Recipient cells	25-cm ² flasks, 70–80% confluent
Ratio of microcells to recipient cells	1 : 1 to 5 : 1
(b) Timetable	
Time	Process
Prior to fusion	Treat plastic bullets used for enucleation with Con A Assemble and sterilize Swinnex filter units
2–3 days prior to fusion	Plate donor cells for micronucleation Plate initially at 25–35% confluence (usually $2.5\text{--}7.5 \times 10^6$ cells per 150-cm ² flask) For 12-bullet enucleation, micronucleate 1–4 150 cm ² flasks (10^6 cells per bullet)
1–2 days prior to fusion	Micronucleate donor cells Add colcemid to donor cell cultures using empirically determined optima of concentration and duration Each 150-cm ² flask should yield about $0.5\text{--}1.5 \times 10^7$ cells
1 day prior to fusion	Plate recipient cells Plate into 25-cm ² flask at a density to be 70–80% confluent at fusion time (usually $1\text{--}3 \times 10^6$ cells per 25-cm ² flask)
Day of fusion	Enucleate micronucleated donor cells Plate 10^6 micronucleated cells per bullet. Enucleate in cytochalasin B Fusion Fuse $0.5\text{--}2 \times 10^7$ microcells to each 25-cm ² flask of recipient cells (microcell-to-recipient cell ratio of 1 : 1 to 5 : 1)
1 day after fusion	Split cells Add selection

line, and differences of 0.01 µg ml⁻¹ colcemid can have a large effect on micronucleation. Table 14.5 lists micronucleation conditions for a variety of cells to serve as a starting point for establishing optimal micronucleation conditions. Most rodent cells micronucleate effectively in colcemid concentrations of 0.01–0.1 µg ml⁻¹ while human cells, especially primary cultures, can require much higher and potentially cytotoxic concentrations in the range of 10–20 µg ml⁻¹. Maximum micronucleation usually occurs at 24–48 h of colcemid exposure.

Rodent cell lines will ordinarily have a micronucleation index of 60–90%. Human cell lines micronucleate less efficiently and it may prove impossible to induce greater than 30% micronucleation. Cytochalasin B treatment during enucleation promotes micronucleation, increasing the yield of microcells. Protocols for cytochalasin B-dependent micronucleation following a mitotic block have been described [100]. Mitotic arrest followed by a hypotonic shock [110] or a cold shock [111] has also been used to micronucleate cells. These methods may be worth trying if prolonged mitotic arrest is unsuccessful.

Protocol 70 gives details of optimization of micronucleation.

14.11.2 Isolation and enucleation of microcells

Microcells are isolated by centrifugation of the micronucleated donor cells in the presence of cytochalasin B [112]. An efficient method for enucleation involves the attachment of the micronucleated cells to a solid surface and centrifugation in cytochalasin B-containing medium. Cytoplasts remain attached to the surface and the microcells pellet. Pretreatment of the surface with concanavalin A (Con A) will firmly attach both adherent and suspension cells and reduces whole cell

contamination. Protocol 71 for the enucleation and isolation of microcells requires the production of ‘plastic bullets’: bullet-shaped pieces of plastic cut from tissue culture plates. These are easily prepared and can be used repeatedly. Methods for enucleation of cells within the flasks in which they are grown have been described, but the flasks are susceptible to breakage and offer no advantage over plastic bullets.

Alternatively, enucleation can be performed through a Ficoll [113] or Percoll gradient [110] in which shear forces due to density differences of the nucleus and cytoplasm fractionate the cells. Both suspension or adherent cells can be enucleated in this manner and large numbers of cells can be processed, but the general efficiency of enucleation is lower than the adhered cell method. Also, the microcell preparation from gradients is crude, and further purification is essential. We suggest using gradient isolation if the adhered method (Protocol 71) proves ineffective. Protocol 72 uses a Percoll isopycnic gradient formed in a centrifugal field [110]. An alternative procedure using a Ficoll step gradient has been described and is also commonly used [113].

14.11.3 Filtration

Karyoplasts will be present in microcell preparations and can represent a large proportion of particles isolated from cells that micronucleate poorly. Membrane filtration (Protocol 73) removes karyoplasts and larger micronuclei, but particles of all sizes are lost in the filtration process, resulting in a reduction in microcells available to fuse. The decision to filter should be based in part on the qualitative assessment of the particles derived from enucleation. If large numbers of karyoplasts are present following enucleation, or if there is no

Table 14.5 Colcemid conditions for micronucleation.

Cell line	Colcemid (µg ml ⁻¹)	Time (h)	Micronucleation (%)	Reference
Primary human fibroblasts	20	48	63	[191]
HT1080-6TG (human fibrosarcoma)	0.01	48	n.r.	[99]
D98/AH2 (HeLa cell)	0.2	48	n.r.	[99]
MEF (mouse embryo fibroblasts)	0.05	36	60–70	[95]
L-M TK ⁻ (mouse L-cell)	0.02	48	90	[96]
A9 (mouse L-cell)	0.1	48	80–90	[95]
L9 (rat myoblasts)	2.0	48	>80	[104]
CHO (Chinese hamster ovary)	0.1	48	n.r.	[192]
Primary chicken fibroblasts	1.0	24	n.r.	[98]
DT40 (chicken pre-B cell)	0.1	24	n.r.	[79]

n.r., Not reported.

selection against the donor cells, the particles should be filtered. Filtration also removes large microcells containing multiple chromosomes, reducing the number of hybrids obtained with large numbers of donor chromosomes. If in doubt, the microcell preparation can be split, fusing one half of the particles following filtration, and fusing the other half unfiltered.

14.11.4 Fusion

A method for the fusion of microcells to whole recipient cells is given in Protocol 74.

14.12 Irradiation and fusion gene transfer (radiation hybrids)

Irradiation and fusion gene transfer (IFGT) is an extension of conventional mammalian cell fusion in which the chromosomes of one parental line (the donor) are fragmented prior to fusion by exposure to ionizing radiation. The hybrids generated in this manner are referred to as radiation hybrids.

There are two common applications for IFGT.

1 *The isolation of hybrids containing small defined regions of specific chromosomes* These 'radiation-reduced' hybrids are usually constructed from a donor cell line that already contains a reduced amount of DNA from the genome being studied, for example, hybrids containing a single human chromosome or subchromosomal fragment [109, 114]. The small amounts of DNA retained in the radiation-reduced hybrids can serve in sublocalization of DNA sequences and as molecular cloning reagents.

2 *The construction of whole chromosome (or whole genome) maps, commonly known as radiation mapping* The probability that two loci will be separated by a radiation-induced break should be proportional to their distance apart: DNA markers close together are likely to be retained or lost together. The further apart two markers are, the more likely it is that a break will be induced in the DNA between them, resulting in their being carried on different fragments. Therefore, the radiation dose determines the extent of fragmentation and in turn the resolution of the map.

Radiation hybrids span the gap between the limits of molecular methods and more conventional somatic cell hybrids, allowing the analysis of DNA fragments of 0.3–30 Mb. An important feature of mapping PCR-based markers using radiation hybrids is that they do not need to be polymorphic

(see below). There has been an exponential increase in the reported use of IFGT in the last 5 years, emphasizing the value of this technology as a tool for mapping chromosomes and deriving new markers from regions of interest. Most IFGT studies have concentrated on the analysis of human chromosomes, although radiation hybrids from other mammalian genomes have been made, for example from the mouse. In conjunction with genetic linkage analysis and physical mapping techniques, IFGT has played an important role in positional cloning experiments designed to isolate human disease genes.

The technique as originally described by Goss and Harris [115–117] used normal diploid cells as the irradiated donor and theoretically could allow the generation of a whole-genome radiation hybrid (RH) map from a single panel of hybrids. However, in recent years most IFGT studies have concentrated on the characterization of panels derived from monochromosomal hybrid donor cells [118]. The feasibility of whole-genome IFGT has recently been re-examined by Goodfellow and colleagues [119]. The results suggested that a single panel of a hundred or so hybrids can be used to map an entire genome. The Whitehead Institute for Biomedical Research/MIT Centre for Genome Research has used a commercially available human whole genome RH panel (Genebridge 4, Research Genetics) consisting of 91 hybrids to map 6193 RH-mapped markers in 23 linkage groups (Release 9, December 1995). A publicly available experimental mapping server allows you to map new STSs screened against the Genebridge 4 panel relative to the Whitehead RH framework map. The RH map and mapping server are accessed via the World Wide Web at http://www-genome.wi.mit.edu/cgi-bin/contig/phys_map (Human Physical Mapping Project section).

Whole-genome IFGT hybrid panels should be extremely powerful tools in developing maps of other species. For example, a whole genome radiation mapping panel has been made for the mouse [120] and is available from Research Genetics (see Appendix III).

Comprehensive reviews of the development of IFGT have recently been published [121–123].

14.12.1 Selection

The dose of radiation used in generating IFGT hybrids is usually lethal to cells, eliminating the need for selection against the donor cells. (Lethal doses differ slightly between cell lines, but most experimenters have found that doses of greater than

1500 rad are sufficient to kill cells.) In order to select for fusion events, the recipient cell line must be deficient in some marker that is present in the donor cells. Selection for a gene function that is present in the genome of the donor cell line, but not from the region of interest (e.g. on a chromosome present in the rodent background of a monochromosomal human/rodent donor cell line), is frequently employed. Two rodent cell lines commonly used as recipients in human gene mapping IFGT experiments are the Chinese hamster derivatives Wg3H (HPRT⁻) and A23 (TK⁻) [36]. These cell lines grow well in culture and exhibit reasonably low reversion frequencies. The choice of parental cell line partners may be important since not all combinations of donor-recipient appear to work in IFGT.

It is not necessary to select for retention of the chromosome to be mapped, since it is usually retained at high frequency in the absence of selection. (IFGT produces hybrid clones containing multiple chromosomal fragments.) However, one can select directly for or against retention of loci on donor chromosomes to enhance the recovery of hybrids that carry the region containing the selectable marker.

14.12.2 Radiation dose

Increasing the radiation dose increases the frequency of chromosome breakage. Siden and co-workers [124] reported that in hybrids generated using a dose of 5000 rad 10% retained entire chromosome arms and 40% had fragments of 3–30 Mb and the remaining 50% retained fragments of less than 2–3 Mb. Using 25000 rad, less than 6% of hybrids had fragments of 3 Mb or larger. It has also been found that increasing the radiation dose increases the percentage of human-positive clones (\approx 50% of hybrids at less than 10 krad contain human material, while nearly 100% of those generated at greater than 10 krad do).

The resolution of the radiation map should therefore be improved by increasing radiation dose, provided the marker retention frequency (the proportion of hybrids which retain a given marker) does not decrease too severely. Radiation hybrids for chromosome mapping are generally generated at low radiation doses (less than 10000 rad); those for positional cloning experiments are generated at high doses. For mapping the important parameter is the number of informative clones; the maximum amount of information is obtained with marker retention frequency of 50% [125]. (For cloning, it is more important that only a few small defined fragments are retained.)

14.12.3 Irradiation and fusion

To establish the order of loci using IFGT hybrids, a panel of radiation hybrids is required (it is not possible to use individual radiation hybrids as mapping tools). Protocol 75 should produce enough hybrids for a panel of mapping hybrids. For *radiation-reduced cloning hybrids*, the procedure is the same, although fewer hybrids are probably required.

Radiation hybrids usually retain multiple fragments of the donor genome (1–10 fragments per cell). Most of the human material is retained by integration into the rodent genome, but some free fragments are found. Radiation hybrids are therefore often unstable, particularly those generated using high radiation doses. Several factors may influence the relative retention of different regions of the chromosome; for example, fragments that already contain functional chromosome elements, such as centromeres, may have a selective advantage. (Markers near centromeres and centromeric alphoid DNA are often retained at higher than average frequencies.) Sequences adjacent to a marker selected for in the fusion will also be retained at high frequency. The number of human fragments retained decreases as the number of cell divisions increases.

Heterogeneous populations within single radiation hybrid cell lines are common. It has been reported that only 10–20% of clones produced using a high dose of irradiation maintained their human DNA component after freezing and thawing. In some cases this inherent instability is useful, as single-cell cloning from existing hybrid lines allows segregation of the various fragments originally retained. Fortunately, the multiple fragments retained within single clones do not appear to undergo extensive rearrangement, that is they maintain their physical integrity, and can be used for short-range or regional mapping and as a cloning resource. They cannot be used for long-range pulsed-field gel electrophoresis mapping.

14.12.4 Analysis of radiation hybrids

The instability of radiation hybrids requires that they must be grown and DNA extracted from a single isolation. Results of marker analysis can only be combined if the same batch of cells has been tested. By far the most efficient method to map large numbers of markers against panels of radiation hybrids is PCR. These PCR-based markers do not need to be polymorphic.

A quick way of estimating the human DNA content of radiation hybrids is by interspersed repetitive sequence PCR (IRS-PCR), either by *PCR*

fingerprinting or by using the IRS-PCR products as chromosome paints in *in situ* hybridization experiments on normal human metaphase chromosomes (see Chapters 9 and 10). IRS-PCR also allows the rapid development of new markers from specific regions of the genome.

Before embarking on large-scale generation, and DNA extraction from, radiation hybrid panels, a sensible precaution is to carry out a simple initial PCR screen of the hybrids for retention of donor material. Once a panel has been generated, such a screen also allows the rapid elimination from further analysis of lines failing to retain human material.

14.12.5 Making the map

RH mapping is a statistical rather than a physical mapping method. Several different mathematical models and methods for the statistical analysis of RH mapping data have been described, each with its own strengths and weaknesses.

With marker retention frequencies between 20 and 50%, analysis of 100 hybrids should provide sufficient data to allow the true marker order to be ascertained. Statistical methods are used to order the markers and estimate the distances between them [118,126–131]. Programs commonly used for RH hybrid mapping include RHMAP and RHMAPPER (see Appendix V for sources). These programs use multipoint mapping analysis based on minimizing obligate chromosome breaks and maximizing the likelihood for several different breakage and retention models. The distances between marker loci are usually expressed in centiRays (cR) and depend on the radiation dose used to generate the hybrids. A distance of 1 cR (centirad) between two markers corresponds to a 1% frequency of breakage between the two markers after irradiation at N rad of X-rays. The estimated relationship between distances in cR and physical distances in kb has been reported as 1 cR = 14.2 kb at 6000 rad [130], 1 cR = 20–75 kb at 6500 rad [133] and 1 cR = 55–75 kb at 9000 rad [134,135].

MultiMap (see Chapter 4) has also been adapted for radiation hybrid mapping.

14.13 Characterization of somatic cell hybrids

The mapping of a phenotype or a DNA sequence using hybrids with a reduced chromosome content requires that the hybrids be as fully characterized for donor DNA content as possible. DNA marker analysis can be used to detect the presence of specific sequences but is not informative about the integrity

of adjacent sequences. Banding and karyotypic analysis may allow positive identification of retained donor chromosomes, but rearranged chromosomes are often difficult to identify and translocations of donor material to the recipient genome are often undetected. Fluorescence *in situ* hybridization (FISH) using IRS painting (e.g. with probes derived from Alu-PCR and DOP-PCR; see below and Chapters 9–11) is very useful in providing information about the donor DNA identity and content in a hybrid. The application of each of the above techniques will depend upon the purpose of a hybrid, and it is necessary to apply them all for full characterization. It is important to be aware that rearrangements/deletions will occur in some hybrids that will not be detected by any method.

14.13.1 Marker analysis

Early somatic cell hybrid experiments utilized enzyme selection or isoenzyme analysis to establish the identity of DNA within the hybrids. This has been largely abandoned in favour of DNA analysis by Southern blot or PCR amplification. PCR is the method of choice, given the relative ease with which large numbers of markers and hybrids can be tested. PCR primer pair sequences are available for all human chromosomes [136–139] and primers can be designed from published sequences that are specific for the genome to be analysed or amplify different-sized products from the background genome.

14.13.1.1 Interspersed repetitive sequence-PCR

Human-specific DNA amplification using primers to interspersed repetitive sequences (IRS) facilitates the characterization of somatic cell hybrids, as well as enabling the rapid development of new markers from specific regions of the genome.

Nelson *et al.* [140] constructed a number of PCR primers using a conserved region of the human Alu repeat sequence. These *short interspersed repeat elements* are present at $\approx 10^6$ copies in the haploid genome of primates. The average density of Alu repeats is one per 4 kb. Variability in distribution of the elements will position some elements within the size range of PCR amplification. The selective amplification of human sequences present in somatic cell hybrids is possible because of the evolutionary divergence of the repetitive elements between man and rodents. The primers do not cross-hybridize with the Alu-related rodent B1 sequence. This makes it possible to isolate DNA markers from specific human chromosomes, or subchromosomal regions, from such hybrids, by specifically amplifying human interAlu sequences. In general, one

Identification of *SOX9* as the gene responsible for both campomelic dysplasia and sex reversal

Mutations in the human *SOX9* gene result in two distinct phenotypes: the congenital skeletal malformations comprising campomelic dysplasia (CD), and in most 46,XY CD patients, male to female sex reversal [205]. *SOX9* was implicated in these phenotypes by association with a translocation breakpoint cloned from a sex-reversed CD patient, with somatic cell hybrids playing a vital role in many aspects of the breakpoint mapping. The original localization of the campomelic dysplasia/sex reversal locus to chromosome 17q24.3-q25.1 was the result of cytogenetic observation of balanced translocations in a number of patients [206,207]. PCR analysis of flow-sorted chromosomes placed the locus between the growth hormone gene (*GH1*) and the thymidine kinase gene (*TK1*) [208], both of which had previously been mapped and sublocalized on chromosome 17 using interspecific somatic cell hybrids containing portions of the human genome.

To identify precisely the position of the translocation breakpoint, it was necessary to construct a high-resolution map of the *GH1-TK1* region of chromosome 17, which was accomplished by using a whole-genome radiation hybrid panel [134]. Again, many of the genes and markers used

in the radiation mapping experiment had been localized to this region of chromosome 17 by hybrid mapping, including one that had been identified as the result of phenotypic studies of microcell hybrids. The ordered markers were then tested on a somatic cell hybrid which was constructed to separate the CD patient translocation chromosome 2pter-q35;17q23-qter from the normal human chromosome 17 and from the reciprocal translocation chromosome [96]. The markers were positioned relative to the breakpoint using this hybrid, as chromosome 17 markers present in the hybrid must be located distal to the breakpoint (i.e. between the breakpoint and the end of the long arm of chromosome 17), while markers not present in the hybrid must be located proximal to the breakpoint. This analysis identified markers close enough together to serve as starting points for a physical contig of sequences.

Finally, the hybrid containing the rearranged chromosome was tested by Southern blotting with subclones of the cosmids for rearranged chromosome 17 sequences. The *SOX9* gene was found to be adjacent to the translocation breakpoint, and single-strand conformation polymorphism (SSCP) mutation analysis (see Chapters 5 and 19) was used to identify *de novo* mutations in sex-reversed CD patients, establishing that alterations in *SOX9* can cause both campomelic dysplasia and autosomal sex reversal [96].

Case Study 14.1

primer complementary to the 3'-end of the Alu repeat consensus is used to amplify the DNA regions flanked by two Alu sequences present in opposite orientation. Because of the presence of a large number of template sites, total human DNA amplified in this way gives a smear, whereas DNA from monochromosomal and subchromosomal hybrids shows closely spaced, defined bands when the PCR products are separated by agarose gel electrophoresis [140].

An inherent bias in the use of Alu-primed PCR is the nonuniform distribution of Alu sequences, which are more frequent in the GC-rich R-bands of chromosomes. The reverse distribution is found for the second major class of interspersed repetitive sequence, the L1 element, a *long interspersed repetitive element* present at 10^4 – 10^5 copies per genome in mammals. This was exploited by Ledbetter *et al.* [141] by constructing species-specific L1 PCR primers. Use of a primer directed against the human L1 sequence produces fewer amplification products than with Alu, consistent with its lower abundance. In principle the use of Alu and L1 primers separately and/or together should optimize genome coverage. Several different types of primer oligonucleotides, based upon the short or long interspersed repeat sequences have been designed for amplification of human DNA present in rodent cells [141, 142].

Individual IRS-PCR products can be purified from

agarose gels, or ligated into plasmid vectors, to screen for single-copy sequences. As an alternative to this multistep process, Monaco *et al.* [143] proposed using Alu-PCR products directly as probes to screen existing libraries, thus eliminating the need for purification, cloning and analysis of each individual Alu-PCR product.

Using human-rodent hybrid cells as a source of RNA, human chromosome-specific cDNA libraries have been made. If total RNA is used as the substrate, cDNA synthesis can be primed with oligonucleotides derived from human Alu sequences thereby constructing a human-specific heteronuclear cDNA library. These hncDNAs will contain exon sequences if the amplified interAlu region contains an exon (which may not always be the case).

IRS-PCR has also been modified to allow the generation of DNA probes from mouse chromosomal fragments present in somatic cell hybrids on either a Chinese hamster background [144] or retained in immortalized human cells [145].

IRS-PCR can also be used to generate chromosome 'painting' probes to be used in FISH (see Chapters 9–11). Hybridizing the IRS-PCR products back to the parental donor cell identifies the donor DNA retained in the hybrid from which the probe was generated [119]. IRS-PCR fingerprinting and painting combine to provide an accurate picture of donor genome content and complexity in a short time.

Protocol 68 Preparation of HAT medium

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.
HAT medium can also be purchased as a concentrated stock (Gibco-BRL).

Materials

- methotrexate (or amethopterin or aminopterin) (Sigma)
- hypoxanthine
- thymidine
- NaOH (1 M)
- HCl (1 M)
- growth medium, e.g. Dulbecco's Modified Eagle's Medium (DMEM) or RPMI (Gibco-BRL)

Method

SOLUTION 1: METHOTREXATE (ALTERNATIVES ARE AMETHOPTERIN OR AMINOPTERIN)

- 1 Add 0.045 g methotrexate to 10 ml distilled H₂O.
- 2 Add 1 M NaOH until the methotrexate dissolves.
- 3 Add 10 ml of distilled H₂O.
- 4 Adjust the pH to between 7.5 and 7.8 with 1 M HCl.
- 5 Make up to 100 ml.
- 6 Filter-sterilize and store at -20 °C.

SOLUTION 2: HYPOXANTHINE AND THYMIDINE (HT)

- 1 Add 0.14 g hypoxanthine to 30 ml distilled H₂O.
- 2 Add 1 M NaOH until the hypoxanthine dissolves.
- 3 Adjust the pH to 10 with 1 M HCl.
- 4 Add 0.039 g thymidine to 35 ml distilled H₂O.
- 5 Combine the hypoxanthine and thymidine solutions and adjust to 100 ml.
- 6 Filter-sterilize and store at -20 °C.

HAT medium is made by adding 1 ml of Solution 1 and 1 ml of Solution 2 to 98 ml of growth medium.

Supplements

BRDU

100×=0.3 g 5-bromo-2'-deoxyuridine per 100 ml H₂O ($\approx 1 \times 10^{-2}$ M, so that 1×= 1×10^{-5} M). Store frozen. Light sensitive.

6-THIOGUANINE (2-AMINO-6-MERCAPTOPURINE)

50×=25 mg in 150 ml H₂O (so that 1×= 2×10^{-5} M). Add 1 N NaOH to dissolve and adjust pH to 9.5 with 1 N acetic acid. Filter-sterilize and store at -20 °C.

8-AZAGUANINE

100×=76 mg in 50 ml (so that 1×= 1×10^{-4} M). Add 1 N NaOH to dissolve; heat to 37 °C if necessary and adjust pH to 9 with 1 N acetic acid.

Protocol 69 Whole-cell fusion of mammalian cells

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Overview

- (a) Monolayer fusion of adherent cells
- (b) Suspension fusion

Materials

- media: normal growth medium as appropriate
- serum-free growth medium as appropriate
- selective medium of choice (see Section 14.9 above)
- polyethylene glycol (PEG) 1500 crystals (NBS Biologicals). Store at room temperature. Alternative: PEG 1500, 50% sterile solution (fusion tested) (Boehringer Mannheim Biochemica). Store at 4 °C protected from light
- 25-cm², 75-cm² culture flasks
- 9-cm tissue culture plates
- centrifuge and conical base centrifuge tubes

- (a) Monolayer fusion of adherent cells

Method

- 1 Plate 1×10^6 cells of each parental cell type together into a 25-cm²

flask the day before fusion. (Cells should be roughly 75–80% confluent at the time of fusion.) Plate 1×10^6 cells of each parental cell into separate 75-cm² flasks to be used as selection controls. Incubate for 16–24 h.

- 2** Prepare 1 per fusion flask of 50% (w/w) PEG 1500 in serum-free medium. Dissolve at 37 °C and filter-sterilize. (The solution is viscous and will be difficult to filter.) Warm to 37 °C before use.
- 3** Fuse one flask at a time as follows. Rinse the flask three times with 5 ml serum-free medium, removing as much medium as possible following the final wash.
- 4** Add 1 ml of the 50% PEG to the inside top of the (inverted) flask.
- 5** Turn the flask over, start timing 60 s and gently rock to spread the viscous solution over the cells.
- 6** At 50 s, tip the flask on end. At 60 s aspirate the PEG from the flask.
- 7** Quickly rinse the cells with 5 ml of serum-free medium. Wash two more times, thoroughly removing the PEG from the cells.
- 8** Add non-selective medium (with serum) and incubate overnight.
- 9** After 24 h split the cells 1 : 10–1 : 20 into selective medium. Add selective medium to the control flasks. Change the medium every 3–5 days to remove dead cells and replenish the selection.

(b) Suspension fusion

Method

- 1** Trypsinize the adherent recipient cell lines and plate 1×10^6 cells into a 75-cm³ flask for a selection control.
- 2** Combine 5×10^6 cells of each parental cell type in a sterile conical base centrifuge tube. Pellet the cells together (1000 g for 10 min), and wash the cells by resuspension in serum-free medium followed by centrifugation.
- 3** Remove the supernatant. Complete removal is important to prevent any dilution of the PEG. Break up the pellet well by flicking the bottom of the tube with your finger.
- 4** Prepare 1 ml per fusion of 50% (w/w) PEG 1500 in serum-free medium. Carefully add 1 ml of prewarmed PEG (37 °C) to the cells, using the tip of the pipette to mix gently as the PEG is slowly added.
- 5** Incubate the cells for 90 s at 37 °C.
- 6** Add 1 ml prewarmed (37 °C) serum-free medium drop-wise, stirring with the pipette tip, over a period of 1 min.
- 7** Add 5 ml of prewarmed (37 °C) serum-free medium drop-wise, stirring with the pipette tip, over a period of 2–3 min.

- 8 Add 10 ml prewarmed (37 °C) medium containing serum, stirring with the pipette tip, over a period of 3 min.
- 9 Centrifuge the cells (1000 g for 10 min), remove the supernatant and gently break up the pellet by tapping the bottom of the tube. Resuspend the cells in growth medium. N.B. Do not pipette up and down to break up the pellet. Plate the cells in non-selective medium into 10 75-cm² tissue culture flasks or into 10 9-cm tissue culture plates. Incubate overnight at 37 °C.
- 10 After 24 h replace the growth medium with selective medium. Add selective medium to the control flasks. Change the medium every 3–5 days to remove dead cells and replenish the selection agent.

.....

Troubleshooting

Too few clones

- *If few clones are obtained from a fusion, it may be possible to obtain sufficient hybrids by simply scaling up the fusion protocol.*
- *If no hybrids are generated, changes in PEG concentration or molecular weight may be necessary (see Section 14.5.3 above).*
- *The use of 10–15% DMSO has been reported to increase fusion efficiencies without changing PEG concentrations [97]. Cellular fusion can be directly monitored by staining the fused cells 3–6 h after fusion with a 10% aqueous Giemsa solution. At this point, at least 10% of the cells should be multinucleate.*
- *Dissimilar cell types may fuse at low efficiencies; using parental cells of a similar type might be necessary.*
- *Varying the parental cell ratios within a range of 1 : 10–10 : 1 may help an inefficient fusion. This is particularly true of fusions with parental cells of significantly different size.*
- *Mycoplasma infection of a culture can effect many cell properties including fusion. All cells should be confirmed to be mycoplasma-free before using in fusion experiments.*

Hybrids with no donor chromosomes

Recipient cells that revert to selection insensitivity give rise to a non-hybrid background in fusion experiments. In the initial evaluation of the success of a fusion, it is important to consider the reversion rate of the parental cells determined by the control selection flasks. Each cell line should be tested for reversion at cell numbers similar to those used in the experiments and reversion controls must be included in each experiment. The morphology of revertants is usually the same as the parental cell and can be useful in distinguishing them from hybrid clones.

.....

Protocol 70 Optimization of micronucleation

Materials

- growth medium as appropriate
- colcemid: (Demecolcine; Sigma). Stock solutions of 1 mg ml⁻¹ in 0.9% (w/v) NaCl are stable for 6 months. Store powder desiccated and protected from light at -20 °C
- Hoechst 33258: Bisbenzimidazole (Sigma). Stock solutions of 50 mg ml⁻¹ are stable indefinitely at 4 °C if protected from light
- 3 : 1 (v/v) methanol : acetic acid
- acetoorcein: 0.5% (w/v) orcein (Sigma) in 50% v/v acetic acid. Store indefinitely at room temperature
- 6-cm tissue culture plates
- 25-cm² culture flasks

Method

- 1 Plate cells into a 6-cm plate or 25-cm² flask at 25–35% confluency (≈ 2.5–7.5 × 10⁵ cells) into non-selective medium. If several cell types are to be tested, the cells can be plated onto glass coverslips and treated together in common petri plates. Prepare enough cultures for several concentrations of colcemid and time points (if cells are going to be stained, see below). Incubate for 16–24 h.
- 2 Add colcemid at several concentrations spanning the expected effective range (refer to Table 14.5).
- 3 Determine the micronucleation index (the percentage of cells that are micronucleate) at 6–8 h intervals, starting at 12 h after colcemid addition and continuing to 48 h. Treatment of cells for greater than 48 h (the length of 1–2 population doublings) significantly lowers transfer frequency. Also check for cytotoxic effects of the colcemid. Micronucleation can be assessed semiquantitatively in most adherent cells by phase-contrast microscopy. For a more accurate quantification, and for suspension cells or adherent cells which do not flatten, monitor micronucleation by fluorescent Hoechst 33258^a or acetoorcein^b staining.

^a Hoechst 33258: Fix pelleted cells in a small volume of 3 : 1 (v/v) methanol : acetic acid. Apply a drop of the fixed cells to a microscope slide and air-dry. Cover the cells with 0.5 mg ml⁻¹ Hoechst 33258 for 1 min, then rinse the slide with water. Add a drop of 50% (v/v) glycerol/PBS, apply a coverslip and visualize under UV illumination (excitation 365 nm, emission 480 nm).

^b Add a drop of cell suspension along with a drop of acetoorcein staining solution to a microscope slide. Apply a coverslip and after 1–2 min visualize under transmitted light.

The yield of cell hybrids decreases with increased exposure to colcemid, so use the concentration that gives the highest micronucleation index with the shortest exposure time. Some cells micronucleate poorly under all conditions, a problem that can be overcome by scaling up the microcell preparation, although in practice this becomes impracticable for cells with micronucleation indices lower than 20%.

.....

Protocol 71 Enucleation from plastic bullets

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Overview

- (a) Construction of plastic bullets
- (b) Concanavalin A crosslinking
- (c) Donor cell preparation
- (d) Enucleation

Materials

- complete growth medium as appropriate
- serum-free growth medium as appropriate
- PBS: 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄, 1.47 mM KH₂PO₄ (pH 7.4). Store at room temperature
- 1% SDS
- concanavalin A (Con A) (Sigma). Prepare solution fresh. Store powder desiccated at 4 °C
- WSC crosslinker: 1-Cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulphonate (Sigma). Prepare solution fresh. Store powder desiccated at -20 °C
- cytochalasin B (Sigma). Stock solutions of 2 mg ml⁻¹ in DMSO are stable indefinitely at 4 °C if protected from light
- acetoorcein (see Protocol 70)
- 95% ethanol
- hot-wire cutter and sandpaper
- 15-cm gridded tissue culture Petri plates
- 150-cm² culture flasks
- 0.2-µm filter
- polycarbonate centrifuge tubes, 50 ml round bottom (Nalgene) with caps
- high-speed refrigerated centrifuge: Sorvall RC-5B with SS-34 or SA-600 rotor, or Beckman J2-21 with JA20 rotor
- haemocytometer

(a) Construction of plastic bullets

Method

- 1** Prepare a cardboard template, ≈ 24 × 86 mm and rounded at one end, that fits comfortably into a 50-ml polycarbonate tube.
- 2** Using the template, trace four bullets on the bottom of a 15-cm tissue culture Petri plate. Typical experiments require 12–24 bullets and although the bullets are re-usable, it is worthwhile preparing several sets. Using gridded Petri plates gives bullets with an orientation: a flat surface to which cells are attached and a bumpy (down) side with moulded grid marks.

- 3** Cut the bullets from the plates with a hot wire cutter and smooth the edges with sandpaper. (A drill-mounted sandpaper wheel facilitates sanding.)
- 4** Check the bullets for fit and adjust if necessary.
- 5** Store the bullets in 95% (v/v) ethanol to keep sterile until needed.

(b) Concanavalin A crosslinking

Sets of bullets crosslinked with Con A can be prepared ahead of time and stored at 4 °C.

- 6** In a laminar flow hood, remove bullets from ethanol and place flat into sterile 15-cm Petri plates (up to five per plate). The bullets should be orientated with their cell attachment side up and edges not touching one another. Ordinarily 12–24 bullets are required per experiment. Allow the ethanol to evaporate from the bullets.
- 7** Prepare solutions:
 - WSC crosslinker (75 mg ml⁻¹ in 0.9% NaCl), 0.6 ml per bullet. Sterilize through a 0.2-μm filter.
 - Con A (15 mg ml⁻¹ in 0.9% NaCl), 0.6 ml per bullet. Con A will not go completely into solution: incubate for 30 min at 37 °C, then filter through Whatman no. 1 paper. Sterilize through a 0.2-μm filter.
- 8** Pipette 0.6 ml of the WSC solution to each dry bullet. Add 0.6 ml Con A solution to the WSC and spread the mixture over the surface and to the edges of each bullet. Allow to sit 1–2 h at room temperature.
- 9** Remove the WSC/Con A solution by aspiration. Flood the bullets with 20 ml PBS to wash. Aspirate and repeat. Store crosslinked bullets in 50 ml tubes in PBS at 4 °C.
- 10** Bullets are re-usable but must be re-crosslinked. Following use, soak the bullets in 1% SDS, rinse well in double-distilled H₂O and store in ethanol.

(c) Donor cell preparation

- 11** Plate donor cells at 25–35% confluency (usually $2.5\text{--}7.5 \times 10^6$ cells per 150 cm² flask). Typically, 8–12 bullets will yield enough particles for a single fusion. Micronucleate 1–4 150-cm² flasks for 12 bullets (10^6 cells per bullet). Each flask should yield about $0.5\text{--}1.5 \times 10^7$ cells following micronucleation.
- 12** Incubate cells 16–24 h before addition of colcemid.
- 13** Treat the cells with colcemid using the predetermined optimal conditions of concentration and time.

- 14 Harvest donor cells, pool and quantify the total number of cells. Resuspend in PBS to a final concentration of 10^6 cells per 1.5 ml (0.67×10^5 cells per millilitre).
- 15 Stain a sample of the cells (see Protocol 70, step 3) and determine the proportion of micronucleate, mononucleate and mitotic cells.

(d) Enucleation

- 16 Remove the crosslinked bullets from PBS and place flat into sterile 15-cm Petri plates (up to five per plate). The bullets should be orientated with their cell attachment side up and edges not touching one another. Overlay each Con A-treated bullet with 1.5 ml of the recipient cell suspension (10^6 cells). Allow the cells to adhere 15 min, then check for attachment using a microscope. Cells plated onto Con A-treated surfaces will adhere quickly and firmly.
- 17 Flood each 15-cm Petri plate with 40 ml complete medium and place in an incubator until the cells have flattened. Most adherent cells will flatten in 1–2 h, suspension cells will not flatten but will become firmly attached.
- 18 Prepare enucleation medium, 45 ml per tube ($10 \mu\text{g ml}^{-1}$ cytochalasin B in serum-free medium from a 2 mg ml^{-1} stock in DMSO). Add ≈ 40 ml to each tube and keep at 37°C .
- 19 Prewarm the centrifuge and rotor. The efficiency of enucleation is highly temperature dependent, with little or no enucleation occurring below 25°C . Before enucleation, warm the empty rotor by spinning at the enucleation speed for 15 min at 34°C .
- 20 After the cells have flattened on the bullets, place one or two bullets (back to back) into each tube. The bullets should be completely immersed in enucleation medium. Centrifuge at $28\,000 g$ for 30 min at 34°C .
- 21 Following centrifugation, remove one bullet and evaluate the extent of enucleation by phase-contrast microscopy. If less than 95% complete, spin the remaining bullets another 30 min. When complete, remove the bullets to distilled water (do not allow to dry before cleaning) and decant the medium from the tubes. The enucleation medium can be used for enucleation from a second set of bullets.
- 22 Break up each pellet in the remaining drop of medium, resuspend in 1 ml serum-free medium and pool the suspensions. Stain a small sample of the particles with acetoorcein (Protocol 70, step 3) and determine the relative proportions of microcells, nuclei, whole cells and cytoplasmic vesicles. Use another small sample to quantify total particles using a haemocytometer.

Protocol 72 Percoll gradient enucleation: an alternative protocol for enucleation

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- see Protocol 71

Additional materials

- Percoll (sterile) (Pharmacia Biotech)

Method

- 1 Micronucleate cells as described in Protocol 70. Prepare up to 10⁸ cells per 20 ml gradient. Six to eight gradients usually provide enough microcells for one fusion. Mix Percoll 1 : 1 with the appropriate cell culture medium containing FCS. Prepare 20 ml for each gradient. Add cytochalasin B to a final concentration of 20 µg ml⁻¹.
- 2 Harvest cells, wash once with PBS and resuspend in a small volume of 1 : 1 Percoll: medium.
- 3 Distribute up to 1 × 10⁸ cells per 50-ml polycarbonate centrifuge tube and bring the total volume to 20 ml final.
- 4 Centrifuge at 19 000 r.p.m. in a Sorvall SS34 rotor (27 000 *g* avg.) for 70 min at 34–37 °C.
- 5 Remove the two visible bands of cells from the tubes using a Pasteur pipette. Add the material from each gradient to 25 ml serum-free medium in a conical tube and pellet by centrifuging at 750–1000 *g* for 10 min. Resuspend the pellet in 50 ml serum-free medium and continue with Protocol 73 (filtration). The material harvested from the gradient contains whole cells, karyoplasts and microcells, so filtration must be performed.

Protocol 73 Filtration of microcells

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- serum-free growth medium as appropriate
- acetoorcein (see Protocol 70)

- 25-mm Swinnex Disc Filter Holders (Millipore)
- 8-µm (Nucleopore) and 5-µm (Nucleopore) polycarbonate membrane filters. Assemble the filters, wrap in foil and autoclave
- haemocytometer

Method

- 1** Resuspend the particles in 20 ml serum-free medium for each 12 bullets enucleated. Dilution is important to minimize clogging of the filters and loss of microcells.
- 2** Filter the suspension through an 8-µm and then through a 5-µm filter. The filters are mounted in Swinnex Disc filter holders attached to a 10-ml syringe with the plunger removed. The suspension is poured into the syringe and gently pushed through with the plunger, using a new filter for each 10 ml of particle suspension.
- 3** Centrifuge the purified microcells at 750–1000 g for 10 min. Resuspend the pellets and pool in a final volume of 1 ml serum-free medium.
- 4** Stain a small sample of the particles with acetoorcein (Protocol 70, step 3) and determine the relative proportions of microcells, nuclei, whole cells and cytoplasmic vesicles. There should be a substantial reduction in the relative proportion of large particles. Use another small sample to quantify total particles using a haemocytometer.

.....

Protocol 74 Fusion of microcells to whole recipient cells

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- general non-selective growth medium (with serum and antibiotics)
- serum-free growth medium
- selective medium of choice (see Section 14.9)
- phytohaemagglutinin: P(PHA-P):Lectin (Sigma)
- 0.2-µm filter
- polyethylene glycol (PEG) 1500 crystals (NBS Biologicals). Store at room temperature. Alternative: PEG 1500, 50% sterile solution (fusion tested) (Boehringer Mannheim Biochemica). Store at 4 °C protected from light
- 25-cm², 75-cm² culture flasks

Method

- 1** The day before fusion, plate the recipient cells into a 25-cm² flask. (Typically, 1–3 × 10⁶ cells.) Prepare several extra flasks in case there is a

high yield of microcells. The culture should be roughly 75–80% confluent at the time of fusion. Plate 1×10^6 of each parental cell into separate 75-cm² flasks to be used as selection controls. Incubate 16–24 h.

- 2 Prepare solutions:
 - PHA-P (100 µg ml⁻¹ in serum-free medium). Sterilize through a 0.2-µm filter.
 - PEG 45–50% (w/w) in serum-free medium. (For determination of concentration, refer to Section 14.5.3.) Dissolve at 37 °C and sterilize through a 0.2-µm filter. The solution will be viscous and difficult to filter. Prepare 1 ml per flask of cells to be fused.
- 3 Adjust the concentration of the microcell suspension with serum-free medium to $0.5\text{--}2 \times 10^7$ particles per millilitre (for a fusion ratio of microcells to recipient cells of between 1 : 1 and 5 : 1). Make sure that the microcells are well dispersed.
- 4 Rinse the recipient cells with 5 ml serum-free medium and remove as much of the medium as possible. Add 1 ml PHA-P solution followed by 1 ml of the microcell suspension. Incubate at 37 °C for 10 min. Check that the microcells have agglutinated to the recipient cells, incubating longer at 37 °C if necessary.
- 5 Fuse one 25-cm² flask at a time as follows: Remove as much of the medium as possible. Add 1 ml of PEG to the inside top of the (inverted) flask. Turn the flask over, start timing 60 s and gently rock to spread the viscous solution over the cells. At 50 s, tip the flask on end. At 60 s, aspirate the PEG from the flask. Quickly rinse the cells with 5 ml serum-free medium. Wash two more times, thoroughly removing the PEG from the cells.
- 6 Add non-selective medium (with serum and antibiotics) and incubate overnight.
- 7 After 24 h split the cells 1 : 10–1 : 20 into selective medium. Add selection to the control flasks. Change the medium every 3–5 days to remove dead cells and to replenish the selection. Hybrid clones should be visible in 2–4 weeks.

.....

Troubleshooting

Low hybrid yield

It is not possible to predict the frequency of hybrids to expect from any given microcell fusion experiment. Careful monitoring of each step of the microcell fusion protocol should identify problems that may result in no or low hybrid yield. As many aspects as possible should be tested prior to undertaking a fusion and it may take several experiments to generate hybrids.

Problems with micronucleation

If a particular cell line is resistant to micronucleation, it may be necessary to use a different cell line as the donor. Human primary cells tend to micronucleate poorly and the micronucleation efficiency is greatly reduced as they approach senescence. Use primary cells from as early a passage as possible.

Mycoplasma infection of a culture can affect many cell properties including micronucleation. All cells should be confirmed to be mycoplasma-free before using in microcell fusion experiments.

Problems with enucleation

The most critical aspect of enucleation is temperature [144]. The actual temperature of the medium in the centrifuge tube within the rotor may vary from the centrifuge chamber setting, so it is worth testing the temperature of a solution in a centrifuge tube following centrifugation. A temperature of 34–37 °C is usually appropriate for efficient enucleation, although precise conditions for greater than 95% enucleation should be empirically determined, using temperatures down to 30 °C.

Problems with fusion

Refer to Section 14.5.3 on fusogens for a discussion of variation in polyethylene glycol conditions.

Mycoplasma infection of a culture can affect many cell properties including fusion. All cells should be confirmed to be mycoplasma-free before using in fusion experiments.

Hybrids with no donor chromosomes

Recipient cells that revert to selection insensitivity can give rise to a non-hybrid background in fusion experiments. It is important to consider the reversion rate of a selectable marker in the recipient cells in relation to the low efficiency of microcell fusion. Each cell line should be tested for reversion at cell numbers used in the experiments and reversion controls must be included in each experiment.

Hybrids with too many donor chromosomes

If no selection is used against the donor cells, they can appear as background in the fusion. These clones are often easily discerned as they have the donor cell morphology. Donor cell contamination occurs during enucleation and may be reduced by prespinning the bullets with adhered donor cells (in serum-free medium, not enucleation medium) to remove loosely attached cells.

Larger microcells contain multiple chromosomes, and filtration removes some of these, reducing the complexity of hybrids obtained. An additional filtration of the microcells through a 3- μ m filter may be useful to remove all but the smallest microcells.

Protocol 75 Production of radiation hybrids

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- growth medium with FCS as appropriate
- 75-cm² culture flasks
- equipment for X-irradiation
- materials for whole-cell fusion suspension (see Protocol 69)

Method

- 1 Harvest 5 × 10⁶ donor cells and resuspend in 10 ml of medium with FCS. Place cells in a 75-cm² tissue culture flasks.
- 2 Expose the cells to irradiation from a calibrated medical or industrial X-ray machine. Use the maximum setting (150 kV, 5 mA in a Torrex X-ray machine, no filters) for the time required to deliver the dose required (the specific machine is not important).
- 3 Harvest recipient cells and fuse with the irradiated donors according to the whole-cell fusion suspension fusion protocol (Protocol 69b).

.....

References

1 Kavanagh, K., Walsh, M. & Whittaker, P.A. (1991) Enhanced intraspecific protoplast fusion in yeast. *FEMS Microbiol. Lett.* **81**, 283–286.

2 Carlson, P.S., Smith, H.H. & Dearing, R.D. (1972) Parasexual interspecific plant hybridization. *Proc. Natl Acad. Sci. USA* **69**, 2292–2294.

3 Barski, G., Sorieul, S. & Cornefert, F. (1960) Production dan des cultures *in vitro* de deux souches cellulaires en association, de cellules de caractere 'hybride'. *C. R. Hebd. Seanc. Acad. Sci.* **251**, 1825–1827.

4 Gourdeau, H. & Fournier, R.E.K. (1990) Genetic analysis of mammalian cell differentiation. *A. Rev. Cell Biol.* **6**, 69–94.

5 Pereira-Smith, O.M. & Smith, J.R. (1988) Genetic analysis of indefinite division in human cells: identification of four complementation groups. *Proc. Natl Acad. Sci. USA* **85**, 6042–6046.

6 Stanbridge, E.J. (1988) Genetic analysis of human malignancy using somatic cell hybrids and monochromosome transfer. *Cancer Surv.* **7**, 317–324.

7 Ichikawa, T., Ichikawa, Y., Dong, J. *et al.* (1992) Localization of metastasis suppressor gene(s) for prostatic cancer to the short arm of human chromosome 11. *Cancer Res.* **52**, 3486–3490.

8 Friedberg, E.C., Henning, K., Lambert, C. *et al.* (1990)

Microcell-mediated chromosome transfer: a strategy for studying the genetics and molecular pathology of human hereditary diseases with abnormal responses to DNA damage. *Basic Life Sci.* **52**, 257–267.

9 Lambert, C., Schultz, R.A., Smith, M. *et al.* (1991) Functional complementation of ataxia-telangiectasia group D (AT-D) cells by microcell-mediated chromosome transfer and mapping of the AT-D locus to the region 11q22–23. *Proc. Natl Acad. Sci. USA* **88**, 5907–5911.

10 Legerski, R.J., Liu, P., Li, L. *et al.* (1994) Assignment of xeroderma pigmentosum group C (XPC) gene to chromosome 3p25. *Genomics* **21**, 266–269.

11 Ruddle, F.H. & Creagen, R.P. (1975) Parasexual approaches to the genetics of man. *Ann. Rev. Genet.* **9**, 407–486.

12 Knowlton, R.G., Cohen-Haguenauer, O., Van Cong, N. *et al.* (1985) A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature* **318**, 380–382.

13 Wainwright, B.J., Scambler, P.J., Schmidtke, J. *et al.* (1985) Localization of cystic fibrosis locus to human chromosome 7cen-q22. *Nature* **318**, 384–385.

14 O'Connell, P., Leach, R.J., Ledbetter, D.H. *et al.* (1989) Fine structure DNA mapping studies of the chromosomal region harboring the genetic defect in neurofibromatosis type I. *Am. J. Hum. Genet.* **44**, 51–57.

- 15 Rommens, J.M., Zengerling, S., Burns, J. *et al.* (1988) Identification and regional localization of DNA markers on chromosome 7 for the cloning of the cystic fibrosis gene. *Am. J. Hum. Genet.* **43**, 645–63.
- 16 The European P.K.D. 1 Consortium. (1994) The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell* **77**, 881–894.
- 17 Aula, P. & Nichols, W. (1967) The cytogenetic effects of mycoplasma in human leucocyte cultures. *J. Cell Physiol.* **70**, 281–290.
- 18 Butler, M. & Leach, R.H. (1964) A mycoplasma that induces acidity and cytopathic effect in tissue culture. *J. Gen. Microbiol.* **34**, 285–294.
- 19 McGarrity, G.J., Phillips, D. & Vaidya, A. (1980) Mycoplasma infection of lymphocyte cultures: infection with *M. salivarium*. *In Vitro Cell. Dev. Biol.* **16**, 346–356.
- 20 Levine, E.L., Thomas, L., McGregor, D. *et al.* (1968) Altered nucleic acid metabolism in human cell cultures infected with mycoplasma. *Proc. Natl Acad. Sci. USA* **60**, 583–589.
- 21 Wise, K.S., Cassell, G.H. & Acton, R.T. (1978) Selective association of murine T lymphoblastoid cell surface alloantigens with *M. hyorhinis*. *Proc. Natl Acad. Sci. USA* **75**, 4479–4483.
- 22 Chen, T.R. (1977) In situ detection of mycoplasma contamination in cell cultures by fluorescent Hoechst 33258 staining. *Exp. Cell Res.* **104**, 255–262.
- 23 Barile, M.F., Bodey, G.P., Snyder, J. *et al.* (1966) Isolation of *Mycoplasma orale* from leukemic bone marrow and blood by direct culture. *J. Natl Cancer Inst.* **36**, 155–159.
- 24 Teyssou, R., Poutiers, F., Saillard, C. *et al.* (1993) Detection of mollicute contamination in cell cultures by 16S rDNA amplification. *Mol. Cell Probes* **7**, 209–216.
- 25 Hopert, A., Uphoff, C.C., Wirth, M. *et al.* (1993) Mycoplasma detection by PCR analysis. *In Vitro Cell. Dev. Biol.* **29**, 819–821.
- 26 van Kuppeveld, F.J., Johansson, K.E., Galama, J.M. *et al.* (1994) Detection of mycoplasma contamination in cell cultures by a mycoplasma group-specific PCR. *Appl. Environ. Microbiol.* **60**, 149–152.
- 27 Pontecorvo, G. (1976) Production of indefinitely multiplying mammalian somatic cell hybrids by means of polyethylene glycol treatment. *Somat. Cell Genet.* **1**, 397–400.
- 28 Davidson, R.L. & Gerald, P.S. (1976) Improved techniques for the induction of mammalian cell hybridization by polyethylene glycol. *Somat. Cell Genet.* **2**, 165–176.
- 29 Davidson, R.L., O'Malley, K.A. & Wheeler, T.B. (1976) Polyethylene-induced mammalian cell hybridization: effect of PEG molecular weight and concentration. *Somat. Cell Genet.* **2**, 271–280.
- 30 Schneiderman, S., Farber, J.L. & Baserga, R. (1979) A simple method for decreasing the toxicity of polyethylene glycol in mammalian cell hybridization. *Somat. Cell Genet.* **5**, 263–269.
- 31 Schwartz, A.G., Cook, P.R. & Harris, H. (1971) Correction of a genetic defect in a mammalian cell. *Nature New Biol.* **230**, 5–8.
- 32 Harris, H. & Watkins, J.F. (1965) Hybrid cells derived from mouse and man: artificial heterokaryons of mammalian cells from different species. *Nature* **205**, 640–646.
- 33 Zepp, H.D., Conover, J.H., Hirschhorn, K. & Hodes, H.L. (1971) Human-mosquito somatic cell hybrids induced by ultraviolet-inactivated Sendai virus. *Nature New Biol.* **229**, 119–121.
- 34 Kitt, S., Dubbs, D.R., Piekarski, L.J. & Hsu, T.C. (1963) Deletion of thymidine kinase activity from L cells resistant to bromodeoxyuridine. *Exp. Cell Res.* **31**, 297–312.
- 35 Littlefield, J.W. (1963) The inosinic acid pyrophosphorylase activity of mouse fibroblasts partially resistant to 8-azaguanine. *Proc. Natl Acad. Sci. USA* **50**, 568–576.
- 36 Westerveld, A., Visser, R.P.L.S., Meera Khan, P. & Bootsma, D. (1971) Loss of human genetic markers in man-Chinese hamster somatic cell hybrids. *Nature New Biol.* **234**, 20–24.
- 37 Weiss, M.C. & Green, H. (1967) Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. *Proc. Natl Acad. Sci. USA* **58**, 1104–1111.
- 38 Ringertz, N.R. & Savage, R.E. (1976) *Cell Hybrids* (Academic Press, San Diego).
- 39 Hooper, M.L. (1985) *Mammalian Cell Genetics* (John Wiley & Sons, New York).
- 40 Szybalski, E.H. & Szybalski, W. (1962) Genetics of human cell lines. IV. DNA-mediated heritable transformation of a biochemical trait. *Proc. Natl Acad. Sci. USA* **48**, 2026–2034.
- 41 Szybalski, W. (1992) Use of the HPRT gene and the HAT selection technique in DNA-mediated transformation of mammalian cells: first steps towards developing hybridoma techniques and gene therapy. *BioEssays* **14**, 495–500.
- 42 Littlefield, J.W. (1964) Selection of hybrids from matings of fibroblasts *in vitro* and their presumed recombinants. *Science* **145**, 709.
- 43 Schafer, A.J., Bulla, G.A. & Fournier, R.E.K. (1994) Tissue-specific extinguisher loci in the murine genome: a screening study based on a rat/mouse microcell hybrid panel. *Somat. Cell Mol. Genet.* **20**, 195–213.
- 44 Kusano, T., Long, C. & Green, H. (1971) A new reduced human-mouse somatic cell hybrid containing the human gene for adenine phosphoribosyltransferase. *Proc. Natl Acad. Sci. USA* **68**, 82–86.
- 45 Chan, T.-S., Long, C. & Green, H. (1975) A human-mouse somatic hybrid line selected for human deoxycytidine deaminase. *Somat. Cell Genet.* **1**, 81–90.
- 46 Liskay, R.M. & Patterson, D. (1979) A selective medium (GAMA) for the isolation of somatic cell hybrids from HPRT- and APRT- mutant cells. *Cytogenet. Cell Genet.* **23**, 61–69.
- 47 Baker, R.M., Brunette, D.M., Mankovitz, R. *et al.* (1974) Ouabain resistant mutants of mouse and hamster cells in culture. *Cell* **1**, 9–21.
- 48 Porteous, D.J., Morten, J.E.N., Cranston, G. *et al.* (1986) Molecular and physical arrangements of human DNA in Hras1 selected, chromosome-mediated gene transfectants. *Mol. Cell Biol.* **6**, 2223–2232.

- 49 Scrambler, P.J., Law, H.-Y., Williamson, R. & Cooper, C.S. (1986) Chromosome mediated gene transfer of six DNA markers linked to the cystic fibrosis locus on human chromosome seven. *Nucleic Acids Res.* **14**, 7159–7174.
- 50 Trent, J.M., Stanbridge, E.J., McBride, H.L. *et al.* (1990) Tumorigenicity in human melanoma cell lines controlled by introduction of human chromosome 6. *Science* **247**, 568–571.
- 51 Yamada, H., Fujimoto, S., Barrett, J.C. & Oshimura, M. (1990) Multiple chromosomes carrying tumor suppressor activity for a uterine endometrial carcinoma cell line identified by microcell-mediated chromosome transfer. *Oncogene* **5**, 1141–1147.
- 52 Pereira-Smith, O.M. & Ning, Y. (1992) Molecular genetic studies of cellular senescence. *Exp. Gerontol.* **27**, 519–522.
- 53 Flejter, W.L., McDaniel, L.D., Askari, M. *et al.* (1992) Characterization of a complex chromosomal rearrangement maps the locus for *in vitro* complementation of xeroderma pigmentosum group D to human chromosome band 19q13. *Genes Chromosomes & Cancer* **5**, 335–342.
- 54 Jeggo, P.A., Hafezparast, M., Thompson, A.F. *et al.* (1992) Localization of a DNA repair gene (XRCC5) involved in double-strand-break rejoining to human chromosome 2. *Proc. Natl Acad. Sci. USA* **89**, 6423–6427.
- 55 Jones, C. (1987) Genetic mapping of human chromosome 11: use of human cell surface antigens. *Somat. Cell Mol. Genet.* **4**, 373–374.
- 56 Rettig, W.J., Grzeschik, K.-H., Yenamandra, A.K. *et al.* (1988) Definition of selectable cell surface markers for human chromosomes & chromosome segments in rodent-human hybrids. *Somat. Cell Mol. Genet.* **14**, 223–231.
- 57 Glaser, T., Houseman, D., Lewis, W.H. *et al.* (1989) A fine-structure deletion map of human chromosome 11p: analysis of J1 series hybrids. *Somat. Cell Mol. Genet.* **15**, 477–501.
- 58 Kao, F.-T. & Puck, T.T. (1968) Genetics of somatic mammalian cells. VIII. Induction and isolation of nutritional mutants in Chinese hamster cells. *Proc. Natl Acad. Sci. USA* **60**, 1275–1281.
- 59 Kao, F.-T., Chasin, L. & Puck, T.T. (1969) Genetics of somatic mammalian cells. X. Complementation analysis of glycine-requiring mutants. *Proc. Natl Acad. Sci. USA* **64**, 1284–1291.
- 60 Kao, F.-T. & Puck, T.T. (1972) Genetics of somatic mammalian cells. XIV. Genetic analysis *in vitro* of auxotrophic mutants. *J. Cell Physiol.* **80**, 41–50.
- 61 Dorin, J.R., Emslie, E., Hanratty, D. *et al.* (1992) Gene targeting for somatic cell manipulation: rapid analysis of reduced chromosome hybrids by Alu-PCR fingerprinting and chromosome painting. *Hum. Mol. Genet.* **1**, 53–59.
- 62 Farr, C., Fantes, J., Goodfellow, P. & Cooke, H. (1991) Functional reintroduction of cloned telomeres into mammalian cells. *Proc. Natl Acad. Sci. USA* **88**, 7006–7010.
- 63 Housman, D. (1987) Insertion of dominant selectable markers into the human genome. *Somat. Cell Mol. Genet.* **13**, 441–446.
- 64 Subramani, S. & Southern, P.J. (1983) Analysis of gene expression using simian virus 40 vectors. *Analyt. Biochem.* **135**, 1–15.
- 65 Kriegler, M. (1990) Assembly of enhancers, promoters and splice signals to control expression of transferred genes. *Meth. Enzymol.* **185**, 512–527.
- 66 Morgenstern, J.P. & Land, H. (1990) A series of mammalian expression vectors and characterization of their expression of a reporter gene in stably and transiently transfected cells. *Nucleic Acids Res.* **18**, 1068.
- 67 Jimenez, A. & Davies, J. (1980) Expression of a transposable antibiotic resistance element in *Saccharomyces*. *Nature* **287**, 869–871.
- 68 Colbere-Garapin, F., Horodniceanu, F., Kourilsky, P. & Garapin, A. (1981) A new dominant selectable marker for higher eukaryotic cells. *J. Mol. Biol.* **150**, 1–14.
- 69 Mansour, S.L., Thomas, K.R. & Capecchi, M.R. (1988) Disruption of the protooncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348–352.
- 70 Southern, P.J. & Berg, P. (1982) Transformation of mammalian cells to antibiotic resistance with a bacterial gene under the control of the SV40 early region promoter. *J. Mol. Appl. Genet.* **1982**, 327–341.
- 71 Cepko, C.L., Roberts, B.L. & Mulligan, R.C. (1984) Construction and application of a highly transmissible murine retrovirus shuttle vector. *Cell* **37**, 1053–1062.
- 72 Mulligan, R.C. & Berg, P. (1980) Expression of a bacterial gene in mammalian cells. *Science* **209**, 1422–1427.
- 73 Mulligan, R.C. & Berg, P. (1981) Selection of animal cells that express the *Escherichia coli* gene coding for xanthine-guanine phosphoribosyltransferase. *Proc. Natl Acad. Sci. USA* **78**, 2072–2076.
- 74 Gritz, L. & Davies, J. (1983) Plasmid-encoded hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* **25**, 179–188.
- 75 de la Luna, S., Soria, I., Pulido, D., Ortin, J. & Jimenez, A. (1988) Efficient transformation of mammalian cells with constructs containing a puromycin-resistant marker. *Gene* **62**, 121–126.
- 76 Morgenstern, J.P. & Land, H. (1990) Advanced mammalian gene transfer: high titer retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.* **18**, 3587–3596.
- 77 Hartman, S.C. & Mulligan, R.C. (1988) Two dominant-acting selectable markers for gene transfer studies in mammalian cells. *Proc. Natl Acad. Sci. USA* **85**, 8047–8051.
- 78 Warburton, D., Gersen, S., Yu, M.T. *et al.* (1990) Monochromosomal rodent-human hybrids from microcell fusion of human lymphoblastoid cells containing an inserted dominant selectable marker. *Genomics* **6**, 358–366.
- 79 Dieken, E.S., Epner, E.M., Fiering, S. *et al.* (1996) Efficient modification of human chromosomal alleles using recombination-proficient chicken/human microcell hybrids. *Nature Genet.* **12**, 174–182.

- 80 Mulsant, P., Gatignol, A., Dalens, M. & Tiraby, G. (1988) Phleomycin resistance as a dominant selectable marker in CHO cells. *Somat. Cell Mol. Genet.* **14**, 243–252.
- 81 Eglitis, M.A. (1991) Positive selectable markers for use with mammalian cells in culture. *Hum. Gene Ther.* **2**, 195–201.
- 82 Cartier, M., Chang, M.W.M. & Stanners, C.P. (1987) Use of the *Escherichia coli* gene for asparagine synthetase as a selective marker in a shuttle vector capable of dominant transfection and amplification in animal cells. *Mol. Cell Biol.* **7**, 1623–1628.
- 83 Simonsen, C.C. & Levinson, A.D. (1984) Isolation and expression of an altered mouse dihydrofolate reductase cDNA. *Proc. Natl Acad. Sci. USA* **80**, 2495–2499.
- 84 Miller, A.D., Law, M. & Verma, I.M. (1985) Generation of helper-free amphotropic retroviruses that transduce a dominant-acting, methotrexate-resistant dihydrofolate reductase gene. *Mol. Cell Biol.* **5**, 431–437.
- 85 O'Hare, K., Benoist, C. & Breathnach, R. (1981) Transformation of mouse fibroblasts to methotrexate resistance by a recombinant plasmid expressing a prokaryotic dihydrofolate reductase. *Proc. Natl Acad. Sci. USA* **78**, 1527–1531.
- 86 Kimura, T., Tatatsuki, A., Yamaguchi, I. (1994) Blastidin S deaminase gene from *Aspergillus terreus* (BSD): a new drug resistance gene for transfection of mammalian cells. *Biochim. Biophys. Acta.* **1219**, 653–659.
- 87 Kaufman, R.J., Murtha, P., Ingolia, D.E., Yeung, C. & Kellems, R.E. (1986) Selection and amplification of heterologous genes encoding adenosine deaminase in mammalian cells. *Proc. Natl Acad. Sci. USA* **83**, 3325–3328.
- 88 Chiang, T. & McConlogue, L. (1988) Amplification and expression of heterologous ornithine decarboxylase in Chinese hamster cells. *Mol. Cell Biol.* **8**, 764–769.
- 89 Wahl, G.M., Padgett, R.A. & Stark, G.R. (1979) Gene amplification causes overproduction of the first three enzymes of UMP synthesis in N-(phosphonacetyl)-L-aspartate-resistant hamster cells. *J. Biol. Chem.* **254**, 8679–8689.
- 90 de Saint-Vincent, B.R., Delbruck, S., Eckhart, W. *et al.* (1981) The cloning and reintroduction into animal cells of a functional CAD gene, a dominant amplifiable marker. *Cell* **27**, 267–277.
- 91 Ueda, K., Cardarelli, C., Gottesman, M.M. & Pastan, I. (1987) Expression of a full-length cDNA for the human 'MDR' gene confers resistance to colchicine, doxorubicin, and vinblastine. *Proc. Natl Acad. Sci. USA* **84**, 3004–3008.
- 92 Mullen, C.A., Kilstrup, M. & Blaese, R.M. (1992) Transfer of the bacterial gene for cytosine deaminase to mammalian cells confers lethal sensitivity to 5-fluorocytosine: a negative selection system. *Proc. Natl Acad. Sci. USA* **89**, 33–37.
- 93 Reid, L.H., Gregg, R.G., Smithies, O. & Koller, B.H. (1990) Regulatory elements in the introns of the human HPRT gene are necessary for its expression in embryonic stem cells. *Proc. Natl Acad. Sci. USA* **87**, 4299–4303.
- 94 Lupton, S.D., Brunton, L.L., Kalberg, V.A. & Overell, R.W. (1991) Dominant positive and negative selection using a hygromycin phosphotransferase-thymidine kinase fusion gene. *Mol. Cell Biol.* **11**, 3374–3378.
- 95 Schwartz, F., Maeda, N., Smithies, O. *et al.* (1991) A dominant positive and negative selectable gene for use in mammalian cells. *Proc. Natl Acad. Sci. USA* **88**, 10414–10420.
- 96 Foster, J.W., Dominguez-Steglich, M.A., Guioli, S. *et al.* (1994) Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* **372**, 525–530.
- 97 Hales, A. (1977) A procedure for the fusion of mammalian cells in suspension by means of polyethylene glycol. *Somat. Cell Genet.* **3**, 227–230.
- 98 Gourdeau, H. & Walker, P.R. (1994) Production of intertypic hybrids between adherent and suspended cell populations. *Biotechniques* **17**, 829–834.
- 99 Fournier, R.E.K. & Ruddle, F.H. (1977) Microcell-mediated transfer of murine chromosomes into mouse, Chinese hamster, and human somatic cells. *Proc. Natl Acad. Sci. USA* **74**, 319–323.
- 100 Fournier, R.E.K. (1981) A general high-efficiency procedure for production of microcell hybrids. *Proc. Natl Acad. Sci. USA* **78**, 6349–6353.
- 101 Killary, A.M. & Fournier, R.E.K. (1995) Microcell fusion. *Meth. Enzymol.* **254**, 133–152.
- 102 Leach, R.J., Thayer, M.J., Schafer, A.J. & Fournier, R.E.K. (1989) Physical mapping of human chromosome 17 using fragment-containing microcell hybrids. *Genomics* **5**, 167–176.
- 103 Whitney, M., Thayer, M., Reifsteets, C. *et al.* (1995) Microcell-mediated chromosome transfer maps the Fanconi anaemia group D gene to chromosome 3p. *Nature Genet.* **11**, 341–343.
- 104 Gourdeau, H. & Walker, P.R. (1994) Evidence for trans-regulation of apoptosis in intertypic somatic cell hybrids. *Mol. Cell Biol.* **14**, 6125–6134.
- 105 Li, Y.K. & Ponce de Leon, F.A. (1992) Partitioning of the chicken genome by microcell hybridization. *Poult. Sci.* **71**, 151–160.
- 106 Tunnacliffe, A., Parkar, M., Povey, S. *et al.* (1983) Integration of *Ecogpt* and SV40 early region sequences into human chromosome 17: a dominant selection system in whole cell and microcell human–mouse hybrids. *EMBO J.* **2**, 1577–1584.
- 107 Saxon, P.J., Srivatsan, E.S., Leipzig, G.V. *et al.* (1985) Selective transfer of individual human chromosomes to recipient cells. *Mol. Cell Biol.* **5**, 140–146.
- 108 Athwal, R.S., Smarsh, M., Searle, B.M. & Deo, S.S. (1985) Integration of a dominant selectable marker into human chromosomes and transfer of marked chromosomes to mouse cells by microcell fusion. *Somat. Cell Mol. Genet.* **11**, 177–187.
- 109 Dowdy, S.F., Scanlon, D.J., Fasching, C.L. *et al.* (1990) Irradiation microcell-mediated chromosome transfer (XMMCT): The generation of specific chromosomal arm deletions. *Genes Chrom. Cancer* **2**, 318–327.
- 110 Sanford, J.A. & Stubblefield, E. (1987) General protocol for microcell-mediated chromosome transfer. *Somat. Cell Mol. Genet.* **13**, 279–284.
- 111 Zatsepina, O.V., Poliakov, V.I. & Chentsov, S. (1982) Nuclear envelope formation around metaphase chromosomes: chromosome decondensation and nuclear

- envelope reconstitution during mitosis. *Eur. J. Cell Biol.* **26**, 277–283.
- 112 Ege, T. & Ringertz, N.R. (1974) Preparation of microcells by enucleation of micronucleate cells. *Exp. Cell Res.* **87**, 378–382.
 - 113 Wigler, M.H. & Weinstein, I.B. (1975) A preparative method for obtaining enucleated mammalian cells. *Biochem. Biophys. Res. Commun.* **63**, 669–674.
 - 114 Graw, S., Davidson, J., Gusella, J. *et al.* (1988) Irradiation-reduced human chromosome 21 hybrids. *Somat. Cell Mol. Genet.* **14**, 233–242.
 - 115 Goss, S.J. & Harris, H. (1975) New method for mapping genes in human chromosomes. *Nature* **255**, 680–684.
 - 116 Goss, S.J. & Harris, H. (1977) Gene transfer by means of cell fusion. I. Statistical mapping of the human X-chromosome by analysis of radiation-induced gene segregation. *J. Cell Sci.* **25**, 17–37.
 - 117 Goss, S.J. & Harris, H. (1977) Gene transfer by means of cell fusion. II. The mapping of 8 loci on human chromosome 1 by statistical analysis of gene assortment in somatic cell hybrids. *J. Cell Sci.* **25**, 39–57.
 - 118 Cox, D.R., Burmeister, M., Price, E.R. *et al.* (1990) Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250**, 245–250.
 - 119 Walter, M.A., Spillett, D.J., Thomas, P. *et al.* (1994) A method for constructing radiation hybrid maps of whole genomes. *Nature Genet.* **7**, 22–28.
 - 120 Schmitt, K., Foster, J.W., Feakes, R.W. *et al.* (1996) Construction of a mouse whole-genome radiation hybrid panel and application to MMUII. *Genomics* **34**, 193–197.
 - 121 Goodfellow, P.N. (1991) Irradiation and fusion gene transfer. In *Methods in Molecular Biology: Gene Transfer and Expression Protocols* (Murray, E.J., ed.), 53–61 (The Humana Press Inc., Clifton, NJ).
 - 122 Walter, M.A. & Goodfellow, P.N. (1993) Radiation hybrids: irradiation and fusion gene transfer. *Trends Genet.* **9**, 352–356.
 - 123 Leach, R. & O'Connell, P. (1995) Mapping of mammalian genomes with radiation (Goss and Harris) hybrids. *Adv. Genet.* **33**, 63–69.
 - 124 Siden, T.S., Kumlien, J., Schwartz, C.E. & Rohme, D. (1992) Radiation fusion hybrids for human chromosomes 3 and X generated at various irradiation doses. *Somat. Cell Mol. Genet.* **18**, 33–44.
 - 125 Lange, K. & Boehnke, M. (1992) Bayesian methods and optimal experimental design for gene mapping by radiation hybrids. *Annu. Hum. Genet.* **56**, 119–144.
 - 126 Boehnke, M., Lange, K. & Cox, D.R. (1991) Statistical methods for multipoint radiation hybrid mapping. *Am. J. Hum. Genet.* **49**, 1174–1188.
 - 127 Chakravarti, A. & Reefer, J. (1992) A theory for radiation hybrid (Goss–Harris) mapping: applications to proximal 21q markers. *Cytogenet. Cell Genet.* **59**, 99–101.
 - 128 Falk, C.T. (1992) Multilocus mapping strategies on chromosome 21 data sets: Comparison of results from family data, radiation hybrids, and somatic cell hybrids. *Cytogenet. Cell Genet.* **59**, 119–121.
 - 129 Green, P. (1992) Construction and comparison of chromosome 21 radiation hybrid and linkage maps using CRI-MAP. *Cytogenet. Cell Genet.* **59**, 122–124.
 - 130 Lawrence, S. & Morton, N.E. (1992) Physical mapping by multiple pairwise analysis. *Cytogenet. Cell Genet.* **59**, 107–109.
 - 131 Weeks, D.E., Lehner, T. & Ott, J. (1992) Preliminary ranking procedures for multilocus ordering based on radiation hybrid data. *Cytogenet. Cell Genet.* **59**, 125–127.
 - 132 Foster, J.W., Schafer, A.J., Critcher, R. *et al.* (1996) A high resolution whole genome radiation hybrid map of human chromosome 17q22–q25.3 across the genes for GH and TK. *Genomics* **33**, 185–192.
 - 133 Warrington, J.A., Bailey, S.K., Armstrong, E. *et al.* (1992) A radiation hybrid map of 18 growth factor receptor, hormone receptor, or neurotransmitter receptor genes on the distal region of the long arm of chromosome 5. *Genomics* **13**, 803–808.
 - 134 Richard, C.W.D., Withers, D.A., Meeker, T.C. *et al.* (1991) A radiation hybrid map of the proximal long arm of human chromosome 11 containing the multiple endocrine neoplasia type 1 (MEN-1) and bcl-1 disease loci. *Am. J. Hum. Genet.* **49**, 1189–1196.
 - 135 Shaw, S.H., Farr, J.E.W., Thiel, B.A. *et al.* (1995) A radiation hybrid map of 95 STSs spanning human chromosome 13q. *Genomics* **27**, 502–510.
 - 136 Abbott, C. & Povey, S. (1991) Development of human chromosome-specific PCR primers for the characterization of somatic cell hybrids. *Genomics* **9**, 73–77.
 - 137 Theune, S., Fung, J., Todd, S. *et al.* (1991) PCR primers for human chromosomes: reagents for the rapid analysis of somatic cell hybrids. *Genomics* **9**, 511–516.
 - 138 Dubois, B.L. & Naylor, S.L. (1993) Characterization of NIGMS human/rodent somatic cell hybrid mapping panel 2 by PCR. *Genomics* **16**, 315–319.
 - 139 Drwinga, H.L., Toji, L.H., Kim, C.H. *et al.* (1993) NIGMS human/rodent mapping panels 1 and 2. *Genomics* **16**, 311–314.
 - 140 Nelson, D.L., Ledbetter, S.A., Corbo, L. *et al.* (1989) Alu polymerase chain reaction: a method for rapid isolation of human specific sequences from complex DNA sources. *Proc. Natl Acad. Sci. USA* **86**, 6686–6690.
 - 141 Ledbetter, S.A., Nelson, D.L., Warren, S.T. & Ledbetter, D.H. (1990) Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. *Genomics* **6**, 475–481.
 - 142 Cotter, F.E., Hampton, G.M., Nasipuri, S. *et al.* (1990) Rapid isolation of human chromosome-specific DNA probes from a somatic cell hybrid. *Genomics* **7**, 257–263.
 - 143 Monaco, A.P., Lam, V.M.S., Zehetner, G. *et al.* (1991) Mapping irradiation hybrids to cosmid and yeast artificial chromosome libraries by direct hybridization of Alu-PCR products. *Nucleic Acids Res.* **19**, 3315–3318.
 - 144 Simmler, M.-C., Cox, R.D. & Avner, P. (1991) Adaptation of the interspersed repetitive sequence polymerase chain reaction to the isolation of mouse DNA probes from somatic cell hybrids on a hamster background. *Genomics* **10**, 770–778.
 - 145 Herman, G.E., Berry, M., Munro, E. *et al.* (1991) The construction of human somatic cell hybrids contain-

- ing portions of the mouse X chromosome and their use to generate DNA probes via interspersed repetitive sequence polymerase chain reaction. *Genomics* **10**, 961–970.
- 146 Shay, J.W. (1987) Cell enucleation, cybrids, reconstituted cells, and nuclear hybrids. *Meth. Enzymol.* **151**, 221–37.
 - 147 Saccone, S., Besati, C., Andreozzi, L. *et al.* (1994) Assignment of the human cytidine deaminase (CDA) gene to chromosome 1 band p35-p36.2. *Genomics* **22**, 661–662.
 - 148 Kelsall, A. & Meuth, M. (1988) Direct selection of Chinese hamster ovary strains deficient in CTP synthetase activity. *Somat. Cell Mol. Genet.* **14**, 149–154.
 - 149 Yamauchi, M., Yamauchi, N., Phear, G. *et al.* (1991) Genomic organization and chromosomal localization of the human CTP synthetase gene (CTPS). *Genomics* **11**, 1088–1096.
 - 150 Mascarello, J.T., Soderberg, K. & Scheffler, I.E. (1980) Assignment of a gene for succinate dehydrogenase to human chromosome 1 by somatic cell hybridization. *Cytogenet. Cell Genet.* **28**, 121–135.
 - 151 Morton, C.C., Eddy, R.L., Shows, T.B. *et al.* (1989) Human AMP deaminase-1 gene (AMPD1) is mapped to chromosome 1. *Cytogenet. Cell Genet.* **51**, 1048.
 - 152 Patterson, D. (1976) Biochemical genetics of Chinese hamster cell mutants with deviant purine metabolism. IV. Isolation of a mutant which accumulates adenylosuccinic acid and succinylaminoimidazole carboxamide ribotide. *Somat. Cell Genet.* **2**, 41–53.
 - 153 Lai, L.-W., Hart, I.M. & Patterson, D. (1991) A gene correcting the defect in the CHO mutant Ade-H, deficient in a branch point enzyme (adenylosuccinate synthetase) of de novo purine biosynthesis, is located on the long arm of chromosome 1. *Genomics* **9**, 322–328.
 - 154 Jones, C., Morse, H.G. & Patterson, D. (1986) Assignment of the gene complementing the auxotrophic marker Ade-F to human chromosome 2. *Am. J. Hum. Genet.* **39**, A158.
 - 155 Patterson, D., Jones, C., Morse, H. *et al.* (1983) Structural gene coding for multifunctional protein carrying orotate phosphoribosyltransferase and OMP decarboxylase activity is located on long arm of human chromosome 3. *Somat. Cell Genet.* **9**, 359–374.
 - 156 Stanley, W. & Chu, E.H.Y. (1978) Assignment of the gene for phosphoribosylpyrophosphate amidotransferase to the pter-q21 region of human chromosome 4. *Cytogenet. Cell Genet.* **22**, 228–231.
 - 157 Barton, J.W., Bleskan, J. & Patterson, D. (1991) Isolation of a human cDNA encoding amidophosphoribosyltransferase and functional complementation of a CHO Ade-A mutant deficient in this activity. *Somat. Cell Mol. Genet.* **17**, 311–322.
 - 158 Barton, J.W., Hart, I.M. & Patterson, D. (1991) Mapping of a locus correcting lack of phosphoribosylaminoimidazole carboxylase activity in Chinese hamster ovary cell Ade-D mutants to human chromosome 4. *Genomics* **9**, 314–321.
 - 159 Dana, S. & Wasmuth, J.J. (1982) Linkage of the leuS, emtB and chr genes on chromosome 5 in humans and expression of human genes encoding protein synthetic components in human–Chinese hamster hybrids. *Somat. Cell Genet.* **8**, 245–264.
 - 160 Urlaub, G. & Chasin, L.A. (1980) Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proc. Natl Acad. Sci. USA* **77**, 4216–4220.
 - 161 Lee, F., Mulligan, R., Berg, P. & Ringold, G. (1981) Glucocorticoids regulate expression of dihydrofolate reductase cDNA in mouse mammary tumour virus chimaeric plasmids. *Nature* **294**, 228–232.
 - 162 Subramani, S., Mulligan, R. & Berg, P. (1981) Expression of the mouse dihydrofolate reductase complementary deoxyribonucleic acid in simian virus 40 vectors. *Mol. Cell Biol.* **5**, 854–864.
 - 163 Creagen, R.P., Chen, S. & Ruddle, F.H. (1975) Genetic analysis of the cell surface association of human chromosome 5 with sensitivity to diphtheria toxin of mouse–human somatic cell hybrids. *Proc. Natl Acad. Sci. USA* **72**, 2237–2241.
 - 164 Hayes, H., Kaneda, Y., Uchida, T. & Okada, Y. (1987) Regional assignment of the gene for diphtheria toxin sensitivity using subchromosomal fragments in microcell hybrids. *Chromosoma* **96**, 26–32.
 - 165 Nakamichi, N.N., Kao, F.-T., Wasmuth, J. & Roufa, D.J. (1986) Ribosomal protein gene sequences map to human chromosomes 5, 8 and 17. *Somat. Cell Mol. Genet.* **12**, 225–236.
 - 166 Dana, S. & Wasmuth, J.J. (1982) Selective linkage disruption in human–Chinese hamster cell hybrids: deletion mapping of the leuS, hexB, emtB and chr genes on human chromosome 5. *Mol. Cell Biol.* **2**, 1220–1228.
 - 167 Arfin, S.M., Cirullo, R.E., Arredondo-Vega, F.X. & Smith, M. (1983) Assignment of structural gene for asparagine synthetase to human chromosome 7. *Somat. Cell Genet.* **9**, 517–531.
 - 168 Smith, M., Arredondo-Vega, F.X., Cirullo, R.E. & Arfin, S.M. (1984) Assignment of the gene for asparagine synthetase to chromosome 7. *Cytogenet. Cell Genet.* **37**, 585–586.
 - 169 Kao, F.-T., Zhang, X., Law, M.L. & Jones, C. (1984) Regional assignment of GLYB (gly-B) to 8q21.1-qter and PGFT (phosphoribosylglycinamide formyltransferase) to 14q22-qter. *Cytogenet. Cell Genet.* **37**, 504–505.
 - 170 Porterfield, B.W., Pomykala, H., Maltepe, E. *et al.* (1993) The use of methylthioadenosine phosphorylase activity to select for human chromosome 9 in interspecies and intraspecies hybrid cells. *Somat. Cell Mol. Genet.* **19**, 469–477.
 - 171 Jones, C., Kao, F.-T. & Taylor, R.T. (1980) Chromosomal assignment of the gene for folylpolyglutamate synthetase to human chromosome 9. *Cytogenet. Cell Genet.* **28**, 181–194.
 - 172 Walter, B., Wagner-McPherson, C., Kwaitkowski, D. *et al.* (1992) Characterization of chromosome 9q-containing hybrids developed from fusion of the GAT hamster cell line with human lymphoblastoid cells. *Annls Hum. Genet.* **56**, 212.
 - 173 Chan, T.-S., Creagen, R.P. & Reardon, M.P. (1978) Adenosine kinase as a new selective marker in somatic cell genetics: isolation of adenosine kinase-deficient mouse cell lines and human–mouse hybrid cell lines containing adenosine kinase. *Somat. Cell Genet.* **4**, 1–12.
 - 174 Jones, C. (1975) Synteny between the pro+ marker and

- human glutamate oxaloacetate transaminase. *Somat. Cell Genet.* **1**, 345–354.
- 175 Jones, C., Wuthier, P., Kao, F.-T. & Puck, T.T. (1972) Genetics of somatic mammalian cells. XV. Evidence for linkage between human genes for lactate dehydrogenase B and serine hydroxymethylase. *J. Cell Physiol.* **80**, 291–298.
 - 176 Kao, F.-T. (1980) Chromosomal assignment of the gene for phosphoribosylformylglycinamide synthetase (PFGS) to human chromosome 14. *J. Cell Biol.* **87**, 291a.
 - 177 Rozen, R., Barton, D., Du, J. *et al.* (1989) Chromosomal localization of the gene for the human trifunctional enzyme, methylenetetrahydrofolate dehydrogenase-methylenetetrahydrofolate cyclohydrolase-formyltetrahydrofolate synthetase. *Am. J. Hum. Genet.* **44**, 781–786.
 - 178 Mascisch, A. & Rozen, R. (1991) Deficient synthesis of MTHFD, a trifunctional folate-dependent enzyme, in the CHO Ade-E mutant. *Somat. Cell Mol. Genet.* **17**, 391–398.
 - 179 Jones, G.E. & Sargent, P.A. (1974) Mutants of cultured Chinese hamster cells deficient in adenine phosphoribosyltransferase. *Cell* **2**, 43–54.
 - 180 Tischfield, J.A. & Ruddle, F.H. (1974) Assignment of the gene for adenine phosphoribosyltransferase to human chromosome 16 by mouse-human somatic cell hybridization. *Proc. Natl Acad. Sci. USA* **71**, 45–49.
 - 181 Ayusama, D., Koyama, H., Iwata, K. & Seno, T. (1981) Selection of mammalian thymidine auxotrophic cell mutants defective in thymidylate synthase by their reduced sensitivity to methotrexate. *Somat. Cell Genet.* **7**, 523–534.
 - 182 Hori, T., Ayusawa, D., Shimizu, K. *et al.* (1985) Assignment of human gene encoding thymidylate synthase to chromosome 18 using interspecific cell hybrids between thymidylate synthase-negative mouse mutant cells and human diploid fibroblasts. *Somat. Cell Mol. Genet.* **11**, 277–284.
 - 183 Nussbaum, R.L., Walmsley, R.M., Lesko, J.G. *et al.* (1985) Thymidylate synthase-deficient Chinese cells: a selection system for human chromosome 18 and experimental system for study of thymidylate synthase regulation and fragile X expression. *Am. J. Hum. Genet.* **37**, 1192–1205.
 - 184 Cirullo, L.R., Arredondo-Vega, F.X., Smith, M. & Wasmuth, J.J. (1983) Isolation and characterization of interspecies heat-resistant hybrids between a temperature-sensitive Chinese hamster cell asparaginyl-tRNA synthetase mutant and normal human leukocytes: Assignment of human asnS gene to chromosome 18. *Somat. Cell Genet.* **9**, 215–233.
 - 185 Sawami, H., Ito, K., Norioka, M. *et al.* (1989) Transfer of the adenosine deaminase (ADA) gene of a B-lymphoblastoid cell line (LCL) by microcell-mediated chromosome transfer technique. *Nippon Ketsueki Gakkai Zasshi* **52**, 1033–1044.
 - 186 Patterson, D., Kao, F.-T. & Puck, T.T. (1974) Genetics of somatic mammalian cells: biochemical genetics of Chinese hamster cell mutants with deviant purine metabolism. *Proc. Natl Acad. Sci. USA* **71**, 2057–2061.
 - 187 Moore, E.E., Jones, C., Kao, F.-T. & Oates, D. (1977) Synteny between glycinamide ribonucleotide synthetase and superoxide dismutase (soluble). *Am. J. Hum. Genet.* **29**, 389–396.
 - 188 Patterson, D., Graw, S. & Jones, C. (1981) Demonstration, by somatic cell genetics, of coordinate regulation of genes for two enzymes of purine synthesis assigned to human chromosome 21. *Proc. Natl Acad. Sci. USA* **78**, 405–409.
 - 189 van Keuren, M.L., Hart, I.M., Kao, F.-T. *et al.* (1987) A somatic cell hybrid with a single human chromosome 22 corrects the defect in the CHO mutant (Ade-I) lacking adenylosuccinase activity. *Cytogenet. Cell Genet.* **44**, 142–147.
 - 190 Brown, C.J. & Willard, H.F. (1989) Noninactivation of a selectable human X-linked gene that complements a murine temperature-sensitive cell cycle defect. *Am. J. Hum. Genet.* **45**, 592–598.
 - 191 Littlefield, J.W. (1966) The use of drug-resistant markers to study the hybridization of mouse fibroblasts. *Exp. Cell Res.* **41**, 190–196.
 - 192 de Saint-Vincent, R. & Buttin, G. (1973) Studies on 1- β -arabino furanosyl-cytosine-resistant mutants of Chinese-hamster fibroblasts. *Eur. J. Biochem.* **37**, 481–488.
 - 193 Dechamps, M., de Saint-Vincent, B.R., Evrard, C. *et al.* (1974) Studies on 1- β -D-arabinofuranosyl-cytosine (Ara-c) resistant mutants of Chinese hamster fibroblasts. II. High resistance to Ara-c as a genetic marker for cellular hybridization. *Exp. Cell Res.* **86**, 269–279.
 - 194 Ning, Y., Lovell, M., Taylor, L. & Pereira-Smith, O.M. (1992) Isolation of monochromosomal hybrids following fusion of human diploid fibroblast-derived microcells with mouse A9 cells. *Cytogenet. Cell Genet.* **60**, 79–80.
 - 195 Lugo, T.G., Handelin, B., Killary, A.M., Housman, D.E. & Fournier, R.E.K. (1987) Isolation of microcell hybrid clones containing retroviral vector insertions into specific human chromosomes. *Mol. Cell Biol.* **7**, 2814–2820.
 - 196 Doucette-Stamm, L.A., Riba, L., Handelin, B. *et al.* (1991) Generation and characterization of irradiation hybrids of human chromosome 4. *Somat. Cell Mol. Genet.* **17**, 471–480.
 - 197 Lee, J.-Y., Koi, M., Stanbridge, E.J. *et al.* (1994) Simple purification of human chromosomes to homogeneity using muntjac hybrid cells. *Nature Genet.* **7**, 29–33.
 - 198 Koi, M., Shimizu, M., Morita, H. *et al.* (1989) Construction of mouse A9 clones containing a single human chromosome tagged with neomycin-resistance gene via microcell fusion. *Jap. J. Cancer Res.* **80**, 413–418.
 - 199 Kaur, G.P. & Athwal, R.S. (1989) Complementation of a DNA repair defect in xeroderma pigmentosum cells by transfer of human chromosome 9. *Proc. Natl Acad. Sci. USA* **86**, 8872–8876.
 - 200 Saxon, P.J., Schultz, R.A., Stanbridge, E.J. & Friedberg, E.C. (1989) Human chromosome 15 confers partial complementation of phenotypes to xeroderma pigmentosum group F cells. *Am. J. Hum. Genet.* **44**, 474–485.
 - 201 Cuthbert, A.P., Trott, D.A., Ekong, R.M. *et al.* (1995) Construction and characterization of a highly stable

- human: rodent monochromosomal hybrid panel for genetic complementation and genome mapping studies. *Cytogenet. Cell Genet.* **71**, 68–76.
- 202 Siden, T.S., Hoglund, M. & Rohme, D. (1990) Monochromosomal mouse microcell hybrids containing inserted selectable neo genes. *Somat. Cell Mol. Genet.* **16**, 425–435.
- 203 McNeill, C.A. & Brown, R.L. (1980) Genetic manipulation by means of microcell-mediated transfer of normal human chromosomes into recipient mouse cells. *Proc. Natl Acad. Sci. USA* **77**, 5394–5398.
- 204 Worton, R., Duff, C. & Flintoff, W. (1981) Microcell mediated cotransfer of genes specifying methotrexate resistance, emetine sensitivity, and chromate sensitivity with Chinese hamster chromosome 2. *Mol. Cell Biol.* **1**, 330–335.
- 205 Schafer, A.J., Dominguez-Steglich, M.A., Guioli, S. *et al.* (1995) The role of *SOX9* in autosomal sex reversal and campomelic dysplasia. *Phil. Trans. R. Soc. Lond.* **B350**, 271–278.
- 206 Maraia, R., Saal, H.M. & Wangsa, D. (1991) A chromosome 17q de novo paracentric inversion in a patient with campomelic dysplasia., case report and etiologic hypothesis. *Clin. Genet.* **39**, 401–408.
- 207 Young, I.D., Zucollo, J.M., Maltby, E.L. & Broderick, N.J. (1992) Campomelic dysplasia associated with a de novo 2q-17q reciprocal translocation. *J. Med. Genet.* **29**, 251–252.
- 208 Tommerup, N., Schempp, W., Meinecke, P. *et al.* (1993) Assignment of an autosomal sex reversal locus (SRA1) and campomelic dysplasia (CMPD1) to 17q24.3-q25.1. *Nature Genet.* **4**, 170–174.

Chapter 15 Long-range physical mapping

Sebastian Meier-Ewert, Leonard Schalkwyk,
Fiona Francis & Hans Lehrach

Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin, Germany

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

15.1	Introduction, 369	15.7.5	Cosmid clone handling, 376
15.2	Using existing library resources, 369	15.8	P1 library construction, 376
15.3	Choice of cloning system, 369	15.8.1	P1 minipreps, 377
15.3.1	YACs, 370	15.8.2	P1 maxipreps, 377
15.3.2	Cosmids, 370	15.8.3	Rescuing ends from P1 clones, 377
15.3.3	P1 clones, 370	15.8.4	Partial digest mapping of P1 clones, 377
15.3.4	Other cloning systems, 371	15.9	Screening by hybridization, 377
15.4	Genomic DNA preparation, 372	15.9.1	YACs, 378
15.5	YAC library construction, 372	15.9.2	P1s, 378
15.6	Use of YACs, 372	15.9.3	Cosmids, 378
15.6.1	YAC library arraying, 372	Protocol 76	
15.6.2	YAC library filter lifts, 373	Preparation of high molecular weight DNA in agarose from human cell lines or mouse spleens, 378	
15.6.3	Screening of YAC libraries, 373	Protocol 77	
15.6.4	Preliminary characterization of YACs after screening, 373	High molecular weight liquid DNA preparation, 380	
15.6.4.1	YAC agarose block preparation, 373	Protocol 78	
15.6.4.2	Determining the sizes of YACs, 373	YAC library construction, 381	
15.6.5	Chromosome walking with YACs, 374	Protocol 79	
15.6.5.1	Partial restriction digest mapping of YACs, 374	Arraying YAC libraries in microtitre plates, 389	
15.6.5.2	Generation of probes from YACs, 374	Protocol 80	
15.6.6	PCR on YACs, 374	YAC library filter lifts, 389	
15.7	Cosmid libraries, 375	Protocol 81	
15.7.1	Host and vector choice, 375	Preparation of YAC agarose blocks, 391	
15.7.2	Preparation of vector DNA, 375		
15.7.3	Preparation of insert DNA, 376		
15.7.4	Choice of host strain, 376		

Protocol 82

Size fractionation of YACs by pulsed-field gel electrophoresis, 393

Protocol 83

Partial restriction digest mapping of YAC DNA, 394

Protocol 84

Generation of end-specific probes from YAC clones, 395

Protocol 85

Cosmid library construction, 397

Protocol 86

P1 library construction, 401

Protocol 87

Extraction of DNA from P1 clones: minipreps, 414

Protocol 88

Extraction of DNA from P1 clones: maxipreps, 415

Protocol 89

Competition of probes to remove repetitive sequences, 416

Protocol 90

Hybridization and washing, 417

References, 418

15.1 Introduction

Long-range mapping is usually the second step in any positional cloning project. It forms the link between the genetic mapping of phenotypes (see Chapters 1–5) and the isolation of candidate genes (see Chapter 17). The ever more rapidly expanding list of cloned genes involved in genetic diseases, disease susceptibility and developmental control processes is testament to the advances that have been achieved over the last few years in positional cloning techniques (for review see ref. 1). Most positional cloning projects start with the isolation of polymorphic DNA markers that are genetically linked to the phenotype under investigation (see Chapter 5). Ideally, markers both proximal and distal to the genetic locus are used to define a candidate region. The size of the candidate region defined by genetic markers depends on the resolution of the genetic map in that area. At present the resolution of the human genetic map is 1 marker per 5 centiMorgans (cM) and on average 1 cM covers \approx 1 Mb of DNA [2]. For the mouse genetic map there are \approx 4 markers per cM and 2 Mb DNA per cM [3].

The optimal link between genetic and physical mapping will be achieved when the resolution of the genetic map is approximately equal to the size of recombinant DNA that can be isolated by cloning systems. Yeast artificial chromosomes (YACs) can carry an insert of average size of more than 500 kb. The current mouse map has, in addition to many mapped mutations and classical markers, 6183 microsatellite markers, which more than saturate the resolution (92 meioses) at which they were mapped. There is also now a 0.1 cM (200 kb) resolution genetic map [4,5]. The human combined genetic, radiation hybrid, and YAC STS-content map now consists of over 15 000 markers. Over 10 000 loci already have been used to identify YACs [6].

Once genomic clones corresponding to the genetic markers have been isolated, the next step is usually the isolation of overlapping clones in order to walk along the chromosome. This walk is often bidirectional until multiple probes have been mapped that allow the orientation of the emerging contig relative to the gene locus. Once the candidate region has been covered in cloned DNA the search for gene sequences begins (see Chapter 17). This part is often the most time-consuming stage of a positional cloning project and in the absence of clear genetic data can require the isolation and characterization of many expressed sequences before a viable candidate gene is found. A variety of approaches can be taken at most steps in a mapping project [7] and a constant

review of progress is required to identify the most suitable strategy.

15.2 Using existing library resources

There are now many clone libraries stored permanently in microwell plates. At the Imperial Cancer Research Fund (ICRF) a reference library system was set up through which it is possible to obtain many clone libraries in the form of filter arrays. This has now been transferred to the Ressourcen Zentrum at the Max-Planck-Institut für Molekulare Genetik, in Berlin, Germany. There are over 70 arrayed libraries available from this site alone, and there are several other sources of libraries. Some of the World Wide Web sites from which library information can be obtained at the time of writing are:

http://www.dhgp.de/main_e.html

<http://www.hgmp.mrc.ac.uk/homepage.html>

<http://www.cephb.fr/HomePage.html>

<http://www-bio.llnl.gov/bbrp/image/image.html>

After identification of a clone by hybridization, one can obtain a culture of the clone for further analysis. The system has the great advantages of eliminating the need for library construction and of collecting data from many different experiments carried out in a large number of laboratories on a common resource, from which all participants can benefit. In view of the time-consuming process of library construction and initial characterization, the use of an arrayed library can be a considerable saving. Before embarking upon a library construction project we strongly recommend the investigation of existing library resources that might be of use.

15.3 Choice of cloning system

For the purposes of long-range mapping four main cloning systems are extensively used and now have well-established protocols. *Somatic cell hybrids*, *YACs*, *cosmids* and *P1 clones* have all proved powerful tools for long-range mapping and have become cornerstones of most positional cloning projects. Somatic cell hybrids are discussed in Chapter 14 and will not be covered here. The most powerful long-range mapping strategies involve a combination of all these cloning systems, the respective strengths of each complementing each other while their weaknesses are compensated for. The analytical power of these cloning systems has recently been extended by the introduction of large cloned genomic DNA fragments into mouse germline cells, allowing the functional study of genes in the

context of at least some of their genomic environment [8–10].

Representation of sequences within any clone library is an important consideration when embarking on a mapping project. Ideally, libraries should be as large as possible in order to maximize the chances of any given sequence of genomic DNA being represented. In practice, however, there are limits to the number of clones that can be handled conveniently. Assuming that the generation of DNA fragments and their propagation in cloning experiments is sequence independent, then the probability of a given sequence being present can be calculated by the following equation:

$$P = 1 - (1 - f)^N$$

where f is the fraction of the genome that the average insert DNA represents and N is the number of clones. Even though cloning is by no means sequence independent, this equation is a useful guide. As a rule, a 3×genome coverage library (a library that includes three copies of the genome) gives a >95% chance of finding any particular single-copy sequence and is therefore considered sufficient for most mapping purposes. Experience from large-scale, genome-wide mapping projects suggests that libraries with far greater coverage are required if large areas need to be completely covered [11,12] (see Chapter 16). It has also become apparent that several different types of libraries complement each other well in large mapping projects [12,13].

15.3.1 YACs

YACs were first developed in 1987 [14] and have become the medium of choice for the initial stages of most long-range mapping projects. Their main advantage lies in their capacity to carry DNA inserts of several hundreds of kilobases (kb) and often in excess of megabases (Mb) [15]. With YACs, chromosome walking over several megabases is feasible and variably segregating genetic markers can often be assembled in one contig. The main disadvantage of the YAC system lies in the high percentage of clones whose insert DNA does not correspond exactly to the source genomic DNA. Such clones are mostly *chimaeric clones*, in which unrelated genomic regions are present in one clone (reviewed in ref. 16), or clones carrying internal deletions of various sizes. These problems can best be addressed by comparing YAC and genomic mapping data, by mapping onto monochromosomal hybrid panels (see Chapter 14), or by the use of fluorescence *in situ* hybridization (FISH) (see Chapter 9).

15.3.2 Cosmids

Cosmids are plasmids of approximately phage λ size, which are introduced into *Escherichia coli* by *in vitro* packaging and infection as defective λ -phage and circularize *in vivo*. There are many vectors to choose from and some well-characterized hosts. Excellent commercial packaging extracts make cloning relatively straightforward. As in YAC and P1 cloning, however, by far the most advantageous library is one that already exists, particularly individual clones have been picked into microtitre plates and gridded onto high density filter arrays.

For mammalian genomes, the number of clones (hundreds of thousands) needed for a high-coverage library of the whole genome is too great to be readily manipulated in this way by current techniques. Instead, for the human genome, many chromosome-specific cosmid libraries have been constructed (see Section 15.2). A particularly valuable use for such libraries is in 'subcloning' YACs covering a region of interest by using the YACs or Alu-polymerase chain reaction (PCR) products (see Chapters 9 and 11 for protocols for Alu-PCR) derived from them as probes on the cosmid grid. This quickly gives a set of cosmid clones corresponding to the region, which are known to derive from the chromosome of interest rather than from other material which might also be contained in a chimaeric YAC. The cosmids also derive more directly from genomic DNA than would be the case if cosmids were constructed from the YACs.

15.3.3 P1 clones

The P1 cloning system has been developed more recently than either cosmids or YACs and protocols are less well established [17]. Using this system, it is possible to clone pieces of DNA which are at least twice the size of a cosmid insert, with an upper size limit of ≈ 95 kb [17–19]. This size limit is determined by the head capacity of the P1 bacteriophage from which the cloning system is derived. The P1 cloning vector contains several phage-derived sequences including a *pac* site, a plasmid replicon, a lytic replicon, and two *loxP* recombination sites. In order for DNA to be packaged into phage heads, a cleavage at the *pac* site by *pacase* enzyme is needed.

Packaged DNA is linear until the phage infects a specific *E. coli* host which expresses the Cre recombinase gene. Circularization is then possible by site-specific recombination between the two *loxP* sites. The plasmid replicon is responsible for maintaining the circularized recombinant clone at one copy per host cell chromosome, and it is thought

that this low copy number may permit the stable propagation of certain DNA sequences which are unstable in other systems. The lytic replicon is normally repressed but may be induced to increase the copy number of the plasmid for DNA preparation. Other features of the vector include a kanamycin-resistance gene, enabling clone selection and recovery, and a *sacB* gene downstream from the cloning site, with an associated *E. coli* promoter upstream. As expression of the *sacB* gene in the presence of sucrose causes cell death, this enables the positive selection of insert-containing clones on kanamycin + sucrose agar plates.

P1 clones are a valuable complement to cosmids and YACs when a clone resource is being established over a region or genome of interest (as evaluated in depth in the *Schizosaccharomyces pombe* mapping project described in Hoheisel *et al.* [12]. Most often, a YAC contig is developed initially, but the molecular analysis of large fragments of linear DNA in yeast cells can be time consuming and problematic because of the relatively long growth times involved and the susceptibility of the DNA to shearing. For these reasons YACs are generally used to isolate cosmid, P1 or λ -clones which are more easily manipulable. Human P1 clones have now been identified in regions not highly represented or absent from cosmid libraries [13]. Stable P1 clones have also been identified in repetitive regions where cosmid clones have been difficult to isolate [20], and similarly in other regions where cosmid walking attempts have failed and YAC instability has been observed [21]. These examples emphasize the effectiveness of P1 libraries used in parallel with other clone libraries for long-range mapping projects. A P1 cloning kit is commercially available (DuPont Merck), but the number of clones required to generate 3 \times coverage mammalian genome libraries can be daunting and well beyond the limit of one cloning kit. Therefore, arrayed or pooled libraries that are already available are strongly recommended [18,22] as a starting point (see Section 15.2).

15.3.4 Other cloning systems

Several other cloning systems have been developed, designed to accommodate pieces of DNA which are larger (and/or more stable) than cosmids or P1s, and which are propagated in *E. coli* for ease of manipulation (cf. YACs). In one such system P1 cloning has been adapted to package recombinant DNA into bacteriophage T4 heads, which can carry more DNA compared with P1 heads [23]. This system has generated clones with 122-kb inserts,

while still retaining the elaborate system of packaging and site-specific recombination inherent in the P1 system.

Another system takes advantage of the low copy number of the *E. coli* F factor, from which cloning vectors have been constructed [24–27]. Replication of the F factor in *E. coli* is strictly controlled, thus reducing the potential for recombination between cloned DNA fragments. Shizuya *et al.* [27] have described a bacterial artificial chromosome (BAC) system in which 300-kb human DNA fragments have been cloned and stably maintained in an F-factor based vector. Libraries with an average insert size of 150 kb can be achieved. The cloning procedure involves pulsed-field gel electrophoresis (PFGE) for size selection of DNA (cf. YAC and P1 cloning), followed by ligation to vector to create circular products which are then electroporated into *E. coli* DH10B (Gibco-BRL) competent cells. These electrocompetent cells permit high efficiencies of electroporation with large plasmids. Disadvantages of this system include lack of positive selection for inserts, and low DNA recovery from the clones. The BAC vector has, however, several desirable features, including T7/SP6 promoters suitable for riboprobe generation, two cloning sites, and several rare-cutter restriction enzyme sites and a *cosN* site, which enable facilitated restriction mapping by partial digestion [28]. The presence of the *cosN* site has also enabled the BAC vector to be used for cloning 40-kb inserts using the cosmid packaging and infection system, and hence offers increased insert stability over conventional multicopy cosmid vectors ('fosmids') [29].

A system has been described which combines several attractive features of the P1 and BAC cloning systems, and this has been named the P1 artificial chromosome (PAC) system [30]. The vector, pCYPAC, retains most of the properties of the P1 cloning vector including the positive selection properties, and the two replicons (plasmid and lytic). However, as in BAC cloning, circular recombinant DNA is electroporated into DH10B competent cells, eliminating the requirement for packaging extracts and *in vivo* site-specific recombination systems. Hence phage head constraints on DNA insert size are also eliminated, and average insert sizes of 130–50 kb have been attained. The PAC system looks the most promising of the more recently developed cloning systems and an arrayed human PAC library suitable for screening by hybridization is already under construction (Pieter de Jong, personal communication).

15.4 Genomic DNA preparation

If there is no arrayed library available, then you will need to isolate genomic DNA for library construction. The isolation of high-quality genomic DNA for library construction is absolutely critical and it is not recommended to commence library construction unless its quality has been verified.

High molecular weight DNA (> 50 kb) is prone to shearing in liquid, and should therefore always be manipulated with care. It is recommended therefore to prepare genomic DNA in agarose for cloning in anything larger than cosmids. In 0.5% agarose, DNA is not subjected to damaging shearing forces, but still remains accessible to proteins for restriction digestion and ligation. This method of DNA preparation has established itself as standard for isolation of high molecular weight DNA.

Protocol 76 describes the isolation of high molecular weight DNA in agarose and Protocol 77 describes the preparation of a high molecular weight liquid DNA.

15.5 YAC library construction

YAC libraries have become a major resource in genome analysis. There is now a wide variety of YAC vectors available and a range of *Saccharomyces cerevisiae* host strains to choose from. In practice, however, the choice is limited by the low efficiencies of cloning associated with many vector and host systems. Most of the large YAC libraries generated to date have been constructed using the originally published pYAC vectors and the host strain AB1380 [14]. For this reason there has been a trend towards the development of YAC postcloning modification vectors that take advantage of the host homologous recombination system to achieve fragmentation [31], and insertion of selectable markers [32] and copy number control elements [33]. Much work has also been directed at improving the host strains used in library construction, combining high transformation efficiencies with low recombination activity to reduce the frequency of chimaeric clones [34]. Recent preliminary reports of recombination-deficient yeast host strains have indicated that the rate of chimaerism can be reduced significantly.

We have used the pCGS966 vector, which, in combination with the AB1380 host, has several advantages over the pYAC vectors. The main advantage of pCGS966 is that it has a different bacterial antibiotic resistance gene in each arm. This means that both YAC ends can be isolated by *plasmid rescue* (see Section 15.6.5.2 and Protocol 84) However, copy number amplification does not work well

in AB1380, because the host is effectively Gal⁻ although this is not shown in its original genotype [14].

The construction of YAC libraries is a nontrivial task involving several poorly understood steps. There are many steps that can be optimized separately and a variety of conditions can lead to successful library construction [35,36]. We have designed a protocol that has consistently yielded a reasonable number of clones using AB1380 as the host strain. The protocol is an adaptation of previously published protocols [37]. Other host/vector combinations have also been used successfully to construct YAC libraries [36]. Transformation is the most variable and inscrutable step in the YAC cloning process. Detailed optimization is slow and difficult and we do not claim the following to be highly optimized. We have adopted a standardized set of conditions that reflect what has been done in previous successful experiments. This standardization extends to details whose importance is unknown to us and probably small, but we feel it is necessary because it gives a small assurance of consistently obtaining some transformants from ligations that are often precious.

Protocol 78 describes the construction of a YAC library.

15.6 Use of YACs

15.6.1 YAC library arraying

Arraying YAC clones into a permanent storage medium is the most efficient way of maintaining the library as a long-term resource and allows the library to be maintained in several copies [38]. The amount of effort invested in generating and analysing the libraries is far greater than the work involved in arraying the clones into storage microwell plates. There are a range of manual and automated devices available for spotting clones arrayed in microwell plates at high density onto membranes for screening purposes.

Yeast colonies can be picked directly from the library plates into liquid media. As clones are grown only on single-selection medium (–uracil), after transformation it is necessary to grow clones under double selection (–uracil, –tryptophan) at some point before screening, in order to select specifically for those clones that contain both vector arms. In our laboratory, clones are picked using a 12-pinned wheel, mounted on a short handle, so that a colony can be picked onto each pin and then inoculated into one row of a microtitre plate. Protocol 79 gives a method for arraying a YAC library.

15.6.2 YAC library filter lifts

An alternative to arraying YAC clones into micro-well plates is to replicate the primary clones onto double-selection (–ura, –trp) agar plates and then take lifts from these plates for screening. The replication of primary clones is necessary for growing YACs under double selection and to get all colonies growing on the surface of the plates so that filter lifts can be taken. Replication of clones is achieved with a 40000 pin replication device described by Larin *et al.* [35] which covers the same area as a NUNC bioassay plate (Protocol 80).

15.6.3 Screening of YAC libraries

YAC libraries are screened mainly in two forms, either by hybridization or by PCR. Hybridization screening allows a wide range of DNAs ranging in complexity to be used as probe in the absence of any sequence information, but does require the generation of hybridization filters. PCR screening, on the other hand, requires pools of YAC DNAs as target and sufficient information about the probe DNA sequence to generate PCR primers. Methods for hybridization screening can be applied to YAC, cosmids and P1 clones and are given in Section 15.9 (Protocols 89 and 90). It should be kept in mind that the quantity of YAC DNA present on the filter is low, and probes should be of high specific activity (aim for $1 \mu\text{Ci ng}^{-1}$), and contain 100 bp or more of unique sequence. Longer fragments, such as entire cosmids [39], are especially good probes.

Screening YAC libraries by PCR of pools requires an initial investment of effort to make pools (if not available collaboratively or commercially), and many PCR reactions per positive recovered. None the less, it has become very important because it only requires access to primers or primer sequences, and it generally gives an unambiguous result. The pooling scheme now considered standard [6,40] involves dividing the library into 'blocks' of eight 96-well plates and making for each block:

- superpools containing the entire 768 YACs;
- eight pools of 96 YACs representing each plate;
- eight pools of 96 YACs each representing a row through the eight plates; and
- 12 pools of 64 YACs representing each column through the plates.

A typical YAC library of 20 000 clones can thus be screened by assaying 25 superpools, and for each positive superpool, the 28 plate, row and column pools.

Recovery of complete YAC addresses would no doubt improve if a pooling scheme with some

redundancy was used. YACs should be pooled after growth, to minimize problems of variable representation. Pools can be prepared from three replicas of the library in 96-well plates containing SD medium, one used for plate pools, one for rows, one for columns. Pooling can be done with a 96-channel pipettor such as the Costar Transtar 96 and disposable row and column 'reservoir liners'. Once pooled, the cells can be processed into agarose blocks as usual. For example, 96 200- μl cultures at a density of $1 \times 10^7 \text{ ml}^{-1}$ can be pooled to give a volume of 19.2 ml of culture. If this is made into an agarose plug (Protocol 81) and, after washing, melted and diluted to 1 ml, each microlitre of the dilution will contain 2×10^3 molecules of each YAC.

15.6.4 Preliminary characterization of YACs after screening

YACs identified during screening should be streaked on selective medium (–uracil, –tryptophan) and allowed to grow for 1–2 days at 30 °C. In case of difficulty in regenerating the clones from frozen stocks, they should be streaked on nonselective YPD agar. When colonies are of sufficient size, filter lifts for secondary screening can be made and processed for hybridization as described above. Single colonies should be inoculated into selective media and grown for 48 h at 30 °C for agarose block DNA preparation. It is advisable to make a large number of agarose blocks from the initial culture, because in some cases subsequent culturing of the clone can result in rearrangements or deletions of the YAC DNA and it is therefore useful to perform all the analysis on DNA derived from one culture. The first step in the analysis of a YAC clone is to size it by PFGE and then to hybridize it with the isolating probe. This not only helps to confirm the correct clone but also reveals any cotransformed or deleted recombinant DNA fragments. When multiple YACs are found in one clone then it is best to retransform the entire block (yeast + YAC DNA) into fresh spheroplasts and re-screen the transformants.

15.6.4.1 YAC agarose block preparation

Protocols differ in the way in which the blocks are processed. The method given here (Protocol 81) uses Novozym 234 to spheroplast the cells and either proteinase K and sarkosyl, or lithium dodecyl sulphate to prepare the DNA.

15.6.4.2 Determining the sizes of YACs

We use the Biorad Clamped Homogeneous Electric Fields (CHEF) system for PFGE. A brief outline of our methodology is given in Protocol 82.

15.6.5 Chromosome walking with YACs

Having isolated a YAC in the region of interest one has to orientate it relative to the centromere, generate new probes for isolating more overlapping YAC clones, and generate polymorphic markers for genetic mapping. It is important to be aware of the problem of chimaerism when using YACs. Chimaeric DNA inserts may arise by a process of homologous recombination in yeast when two or more DNA fragments are cotransformed into the host yeast strain [16], or may be due to coligation. Consequently, it is useful to have several YACs from a region and to map any probes isolated using genetics and somatic cell hybrid mapping panels.

15.6.5.1 Partial restriction digest mapping of YACs

Partial restriction digest mapping of YACs with rare-cutting restriction enzymes allows correlation with the physical map constructed from genomic DNA, and allows probes to be physically mapped onto the YAC. Yeast DNA is unmethylated and more sites are therefore cut in YAC DNA than in methylated genomic DNA isolated from tissues, but it is still possible to make correlations with genomic DNA. The pYAC4 vector contains different parts of pBR322 in each of its arms; the *right* arm (URA3) can be detected with a 1.4-kb DNA fragment of pBR322 from the *PvuII-SalI* digest and the *left* arm (TRP, AMP^r) can be detected with the 2.3-kb DNA fragment of pBR322 from the *PvuII-EcoRI* digest [14].

Any rare-cutting restriction enzyme can be used, but a certain amount of optimization must be carried out for each enzyme as well as for each batch. Protocol 83 for partial restriction digest mapping of YAC DNA uses *BssHII*.

15.6.5.2 Generation of probes from YACs

One of the most important aims in any long-range mapping project is to isolate probes from the ends of the YAC in order to identify overlapping clones. The success of any walking experiment is crucially dependent on the generation of useful probes from existing clones. Methods for generating both end-specific (Protocol 84) and random probes (Section 15.6.6) are given.

It is worth noting at this point that probes often require competition with genomic DNA in order to suppress the hybridization of repetitive elements (see Protocol 89).

End-specific probes are isolated either by plasmid rescue (Protocol 84a) or as vectorette probes (Protocol 84b). The pYAC4 vector has the pBR322 origin of replication and ampicillin resistance gene

in its *left* arm [14]. The telomere sequence can be digested off this arm by *XhoI* digestion and, providing that there is another such site in the insert within a reasonable distance of the vector, a plasmid can be 'rescued' by ligating the ends of the DNA fragment carrying the vector and some insert DNA together and transfecting it into a bacterial host.

The alternative and effective vectorette probe method [41] uses a linker cassette ligatable to a number of frequently cutting restriction enzyme sites. A YAC block is digested and ligated with such a linker and then vector-specific primers are used in PCR reactions with a primer specific to the linker. The design of the linker is such that priming from the linker cannot occur until the linker is copied by a vector-primed strand, ensuring that only the vector ends are amplified. The main advantage of this method is that both ends of the YAC can be obtained.

15.6.6 PCR on YACs

Interspersed repeat sequence PCR (IRS-PCR, e.g. Alu-PCR on clones of human DNA, B1-PCR on mouse DNA) has proved a very convenient way of recovering a dispersed subset of insert sequences from YAC clones. The PCR products are suitable for use as probes to isolate cosmid or P1 clones corresponding to the YAC from an appropriate library. This is by far the easiest way of 'subcloning' a YAC if a suitable gridded cosmid or P1 library is available. The PCR products from a YAC can also be used singly or as a pool as walking probes for isolating further YACs. The latter is particularly useful when the hybridization targets are themselves PCR products. It is possible to use PCR to amplify entire YAC libraries as individual clones and spot the products on high-density gridded filters [42]. This allows overlaps between YAC clones to be discovered by a technically easy low-complexity hybridization.

Agarose block preparations of YAC DNA are a very good substrate for PCR, and are often available as they are made for a variety of other purposes. A typical block containing cells from 1 to 2 ml of culture can be washed in water, melted at 68°C in 1 ml water and 1 µl used in a PCR reaction.

Alternatively, DNA preparations can be made very quickly from large numbers of YACs in microtitre dishes by the method of Chumakov [43]. This works equally well with the substitution of Novozym 234 (8 mg ml⁻¹) for Zymolyase and can be readily adapted to quadruple-density microtitre dishes (Genetix, Dorset UK). The resulting preparations are very crude and dilute but nonetheless reliably give rise to PCR products. Approximately

100 nl (the amount that adheres to a multipin transfer device or the outside of a pipette tip dipped into the solution) is used in a PCR reaction.

Typical reaction conditions (30- μ l reaction) are:

- 1 μ g of repeat primer (1 μ g each if two primers are used),
 - 1 U Taq polymerase,
- in:
- 200 μ M each dNTP;
 - 50 mM KCl;
 - 1.5 mM MgCl₂;
 - 35 mM Tris base;
 - 15 mM Tris-HCl pH 5;
 - 0.1% Tween-20.

Amplification is for 35 cycles with annealing temperature depending on the primers used. (See also Chapter 9, Protocol 43 for an alternative methodology using purified YAC DNA.)

PCR products can be efficiently random-prime labelled [44] by simply using a sufficiently small volume of the unpurified PCR product so that unlabelled nucleotide carry over is negligible. A typical such reaction uses 5 μ l of a boiled 1:20 dilution in water of PCR products, added to the usual mix of oligonucleotide primers, buffer, pol I Klenow, unlabelled nucleotides and [α -³²P]dATP, [α -³²P]dCTP or both; 20–60 μ Ci are routinely incorporated in such a reaction.

15.7 Cosmid libraries

Protocol 85 describes the construction of a cosmid library.

15.7.1 Host and vector choice

If making a library cannot be avoided, the first job is to choose a vector and host. The essentials of a cosmid vector are a cloning site, a replication origin, a selectable marker and a cos site. In each of these there are some choices, and there are also more specialized vectors containing additional sequences. The cloning site is conventionally a *Bam*H1 site into which *Sau*3A partial digests can be cloned. Other sites, or in a few cases a polylinker [45], are available in some vectors. Some vectors have very useful features flanking the cloning site, such as sites for rare-cutting restriction enzymes and bacteriophage T3, T7, or Sp6 promoters for generating end probes. LoristX and its successors the Lawrist cosmids also have *E. coli* transcription terminators flanking the cloning site in an attempt to reduce possible effects of transcription from the insert on plasmid maintenance.

The most common selectable markers used in

cosmid cloning are the familiar ampicillin resistance (*amp*^R) (mediated by the β -lactamase gene *bla*) or kanamycin resistance (*kan*^R) (mediated by the neomycin phosphotransferase gene, *neo*) which are used in many other plasmid constructs. The *kan*^R marker has several practical advantages. The antibiotic is very stable so media do not have to be prepared freshly as is the case with ampicillin, nor do kanamycin-resistant colonies allow the growth of non-recombinant satellite colonies, as the antibiotic is not metabolized.

15.7.2 Preparation of vector DNA

The first generation of cosmid vectors had pBR322-derived (i.e. pSC101) replication origins. Other plasmid replication origins, such as R6K [46] have been incorporated into cosmids, but most currently used vectors use the λ -replication origin [47,48], which is likely to discriminate less between small and large constructs, and thus reduce the selection for small (and possibly deleted) inserts that has been a problem in cosmid cloning. The use of a λ -origin and the *kan*^R marker also make it possible to produce vectors entirely free of sequences which hybridize with the common pBR322-based plasmid cloning vectors, thus simplifying probe preparation for hybridization to other clones (e.g. P1, YACs or cDNAs).

Early cosmid library construction methods involved ligation of linearized cosmid vector with insert. Both circular and concatameric products containing a complete copy of the vector and a suitable length of insert between two cos sites could be packaged. A more efficient strategy is to make arms, blunt or dephosphorylated at their outer ends. This involves separately preparing two overlapping fragments each containing a copy of the cos site near one end and the cloning site at the other. This has been facilitated by the advent of double-cos vectors which need only be linearized between the cos sites, dephosphorylated and cut with the cloning enzyme [49]. In the case of the Lawrist cosmids [50], the sequence between the cos sites is a pUC series plasmid that allows the vector to be prepared easily in high yield. Just as in YAC cloning, it is a false economy to begin with anything but very pure, tested, vector DNA. We therefore recommend that vector DNA be isolated on a large scale by an alkaline lysis procedure and purified over a CsCl gradient [51]. Protocols for the construction of cosmid and phage libraries and considerations in choice of cloning systems are detailed in [52] and many of the steps in Protocol 85 derive from these protocols.

15.7.3 Preparation of insert DNA

Genomic DNA for use in cosmid cloning can be isolated as described in this chapter (Protocols 76 and 77). For cosmid cloning it is not strictly necessary to use DNA in agarose blocks. Size fractionation of digested DNA can increase the average size of inserts from ≈ 35 kb to closer to 40 kb, but as much larger insert cloning systems are now available (e.g. P1 and YAC) there is less need to generate the largest possible cosmid clones. A size fraction of a partial digest can be cut from a pulsed-field gel as is done for YAC and P1 cloning. Where starting material is limited (e.g. flow-sorted chromosomes, see Chapter 12, or gel-isolated YAC DNA), the best approach is to ligate unfractionated, partially digested, dephosphorylated insert DNA with vector arms. Partial digestions produced by the methylase-competition approach [53] should perform in a nearly concentration-independent manner, and can thus be optimized using a non-precious test material. There are three options for controlling partial digestion of genomic DNA: enzyme concentration, methylation of restriction sites, or time-course assay with a constant amount of enzyme. As the first two systems are described in essence in the P1 and YAC sections of this chapter (Protocols 86 and 78b, respectively) we shall describe the time-course method in Protocol 85.

15.7.4 Choice of host strain

Host strains for cosmid cloning should have the virtues of promiscuity and conservatism: that is to say, they should not discriminate between inserts and should maintain them as stably as possible. The main agents of discrimination are restriction enzymes, of which wild-type *E. coli* strains contain several. In particular, *McrA*, *McrBC* and *Mrr* cut certain sites in C-methylated DNA, such as that from mammals. *EcoK* cuts unmodified *EcoK* sites. It is important that the host strain chosen for library construction (and for packaging extracts) be free of these activities. Examples of good hosts include strains ED8767, DH10B, and DH5 α MCR. Stability of inserts which may contain repeated sequences requires a recombination-deficient host. Because a complete absence of recombination is likely to be incompatible with plasmid maintenance and viability of the host, this will always be something of a compromise. The single mutation which makes the greatest difference in recombination is *recA*, and this is the standard in cloning hosts, including those listed above. Many other recombination-related mutations are available, which are often lethal with

recA. Strains lower in recombination than *recA* strains can be constructed by combining mutations other than *recA*—for example, the commercial SURE strain (Stratagene), which is *uvrC*, *umuC*, *sbcC*, *recB*. For library construction this strain has the important disadvantage of being kanamycin resistant, as well as being a poor grower.

15.7.5 Cosmid clone handling

Clones from cosmid libraries can be treated in many respects like other plasmid clones. DNA can be prepared using the standard alkaline lysis procedure (see ref. 54; see also Chapter 21, Protocol 100) without special precautions, though preparations from non-*endA* hosts should be phenol-extracted. Yields are good, quite consistent in the case of Lorist 2 and its derivatives [55].

If cosmids are to be picked into microtitre plates for storage, this should be done within a few days of plating. The wells should be filled with 2 \times YT medium (per litre: 16 g yeast extract, 10 g tryptone and 5 g NaCl) supplemented after autoclaving with the appropriate antibiotic and with Hogness modified freezing medium (10 \times HMFM contains: 63 g l⁻¹ K₂HPO₄, 18 g l⁻¹ KH₂PO₄, 4.5 g l⁻¹ sodium citrate, 9 g l⁻¹ ammonium sulphate, 440 g l⁻¹ glycerol, and 0.9 g l⁻¹ MgSO₄·7H₂O. The last ingredient should be autoclaved separately in a tenth of the final volume and added when cool). The cultures are allowed to grow to saturation in most wells (overnight at 37°C and stored frozen at -70°C). Cultures are stable indefinitely (as far as can be determined) when stored in this way and multiple rounds of freezing and thawing are possible without significant reduction in viability. We do not recommend -20°C for long-term storage.

Clones identified by screening can be picked either from the primary plate, a frozen lift, or a microtitre well. The best secondary screening method (unless the number to be checked is high) is to prepare a Southern transfer of digested cosmid DNA. Minipreps can be made by the standard alkaline lysis method. With the Lorist or Lawrist series cosmids, the yield from each millilitre of culture in 2 \times YT medium should be easily enough for four gel tracks.

15.8 P1 library construction

Protocol 86 describes the preparation of P1 libraries using DNA embedded in agarose blocks, and performing size selections using PFGE. Some of these methods are similar to those for YAC library construction. Alternative protocols using sucrose

gradient fractionation have also been described [17].

Partial digestions to prepare the insert DNA are considered ideal if a good proportion of the digested DNA gives maximal ethidium bromide fluorescence between 100 and 150 kb on a pulsed-field gel, without appreciable overdigested or underdigested material. Enzymes that produce insert ends compatible with the cloning site in the vector (*Bam*HI) are *Sau*3 A, *Bam*HI, and *Mbo*I. There are several ways in which reactions may be controlled and optimal partial digests may be obtained—for example, using a combination of *Mbo*I and *dam* methylase [53,56], using limiting concentrations of Mg^{2+} [57], different enzyme concentrations and digesting for a fixed time, or performing a time course of reactions with fixed enzyme concentrations.

15.8.1 P1 minipreps

Generally, DNA extraction from P1 clones is not as straightforward as from plasmids and yields less DNA; details of a modified procedure are given in Protocol 87.

Minipreps are performed by the alkaline lysis procedure (modification of ref. 54), starting with a relatively large volume of overnight culture, and hence scaled-up alkaline lysis solution volumes. Clones should be processed individually through the steps involving alkaline lysis buffers (without stopping between each step), before moving on to the next clone (N. Sternberg, personal communication).

Protocol 87 produces fairly crude DNA suitable for sizing undigested DNA on a pulsed-field gel (although this requires marker clones of known size), and for some enzyme digestions. However, it is advisable to incorporate a phenol/chloroform extraction step to obtain a cleaner preparation of DNA (e.g. for rare-cutter enzyme digestions). Clones can be sized by a *Not*I digest (one site in vector) and PFGE, or by using a more frequent cutter and standard gel electrophoresis with Southern blotting.

15.8.2 P1 maxipreps

As for minipreps, it is recommended that all clones are processed through the alkaline lysis solutions I, II and III individually. Protocol 88 describes the extraction of DNA from a P1 maxiprep.

15.8.3 Rescuing ends from P1 clones

Several methods are available for generating end probes from P1 clones.

One approach is derived from a method used to

generate end-rescue probes from YAC clones—the vectorette method [41] (see Protocol 84). This method has been used in an analogous way to generate ends from P1 clones which were prepared in an earlier generation P1 vector [19]. The more recent positive-selection P1 vector, pAd10sacBII, has the advantage of T7 and SP6 promoters flanking the cloning site [58] which are convenient for generating RNA probes (riboprobes). T7 and SP6 sequences can also be utilized in a PCR approach (analogous to the vectorette procedure) whereby an adaptor is ligated to digested P1 clone DNA, and PCR is performed between either the T7 or SP6 promoter and a primer specific for the adaptor [59]. An alternative PCR approach for generating end probes involves primer extension from a radiolabelled vector oligonucleotide (as described for cosmid clones by Hoheisel *et al.* [60]). This method has been used successfully for P1 clones [13], and involves a linear PCR starting with an oligomer which has been end-labelled using [γ^{32} P]ATP and T4 polynucleotide kinase. The PCR product can be used to directly screen library filters or Southern blots after competition with sheared human placental DNA and vector DNA.

15.8.4 Partial digest mapping of P1 clones

The protocols currently available for producing high-resolution restriction maps of P1 clones are ultimately dependent on the rare-cutter enzyme sites present in the insert DNA. The vector DNA sequences flanking the cloning site contain a *Not*I site on one side, and *Sal*I and *Sfi*I on the other. Therefore the absence of one or more of these sites in the insert makes linearization and/or isolation of insert DNA possible. Partial maps of cosmid clones can be created by linearization with terminase at the cos site, partial digestion, and annealing of labelled oligonucleotides complementary to the single-stranded DNA at the terminase-digested cohesive ends [28,61]. A modification of this approach may still be used for mapping P1 clones despite the absence of a cos or similar site in the vector. This procedure requires the identification of enzymes which do not cut in the insert but do cut in the vector (for linearization), and also of specific vector fragments adjacent to the cloning site, to be used as hybridization probes on Southern blots of the linearized and partially digested DNA [13].

15.9 Screening by hybridization

Screening by hybridization involves the annealing of labelled probe DNA to immobilized target DNA in solution usually under non-stringent conditions.

Sequence-specific hybridization is then discriminated from other interactions by a series of washing steps at an increased stringency. The probe DNA can be labelled and detected by a number of systems such as chemiluminescence, fluorescence or radiolabelling. Hybridization protocols vary in detail and there are again several options. We will describe a commonly used method for screening *in situ* DNA filters based on that of Church and Gilbert [62] (Protocol 90). Radiolabelling is still the most commonly used detection system and the following sections are based on this.

15.9.1 YACs

Processed filters bearing spotted or lifted YAC-containing yeast colonies (most commonly 22×22 cM format) can be hybridized in bags by the conventional method (see, for example, ref. 51). Some care is needed because the quantity of YAC DNA present on the filter is small, and some of it is probably retained by cell debris rather than being bound to the filter itself. For this reason gentle treatment of filters is recommended; in particular stripping should be avoided as much as possible. The sensitivity requirement is equal or greater to that for mammalian single-copy genomic Southern blots. Probes should thus have a specific activity of

the order of 0.5 mCi µg⁻¹. DNA is labelled by random priming [44], often using both [α-³²P]dCTP and [α-³²P]dATP. Ease of use increases with length of probe, with fragments below 200bp giving particular difficulty. With all but the best characterized unique sequence probes, competition with total or Cot DNA is essential to remove repeat sequences (Protocol 89).

15.9.2 P1s

Clone lifts, or robotically spotted clone arrays [63], can be processed according to Sambrook *et al.* [51]. However, it is best to incorporate either a wiping step in 2×SSC, or steaming and proteinase K treatment step [56], to reduce nonspecific background hybridization. Probes can also be competed with P1 vector DNA added to a concentration of 25 µg µl⁻¹ to reduce background hybridization.

15.9.3 Cosmids

Cosmid filters are the easiest of the three systems mentioned here to screen successfully. They have the most favourable ratio of clone to host DNA and also the shortest inserts, meaning that nonspecific and repetitive hybridization signal is low. No special precautions are really necessary.

.....

Protocol 76

Preparation of high molecular weight DNA in agarose from human cell lines or mouse spleens

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- 10⁹ cells (or, for example, six mouse spleens) to prepare about 200 agarose blocks (of ≈ 3 × 10⁶ cells per block). It may be preferable to prepare less concentrated blocks for YAC cloning of 1 × 10⁶ cells per block
- 1×TE: 10 mM Tris-HCl, pH 7.5, 1 mM EDTA
- PBSA
- low-melting-point agarose (SeaPlaque CTG, FMC Bioproducts)
- trypsin-versene
- sarkosyl
- proteinase K (BDH)
- phenylmethylsulphonyl fluoride (PMSF) (*Caution: extremely toxic*)

- block formers (Bio-Rad)
- 50-ml Falcon tubes
- centrifuge (Beckman J6B)

Method

- 1** Prepare 1.5% low-melting-point agarose in PBSA, and keep molten at 50 °C. Wash and dry block formers and chill on ice. Block formers consist of 7 × 7 × 2 mm slots in plastic moulds. Block formers are sealed at one side with masking tape to create wells of ≈ 100 µl. Place the sealed block formers on a glass plate sitting on ice to facilitate setting of the agarose when added.
- 2** From cell lines:
 - (a) cell suspension: divide into 50-ml aliquots in Falcon 2098 tubes, or
 - (b) monolayer culture: wash cells twice with PBSA, use trypsin-versene to remove cells from flasks, and transfer to 50-ml Falcon tubes, orfrom mouse spleens:
 - (c) place fresh mouse spleens in a dounce homogenizer with about 20 ml cold PBSA and homogenize carefully until membranes appear clear, but not too vigorously so that cells rupture.Pour contents into a 50-ml Falcon tube (on ice), top up with fresh cold PBSA and allow membranes and debris to settle to the bottom of the tube. Transfer supernatant to fresh 50-ml tube.
- 3** Pellet cells in a Beckmann (J6B) centrifuge for 10 min at 1000 r.p.m.
- 4** Wash twice with 50 ml PBSA, resuspending cells each time.
- 5** Resuspend cells in ≈ 10 ml PBSA total.
- 6** Count cells in a 30-fold and 60-fold dilution. For cells derived from mouse spleen, count all cells then deduct 40% to account for the red blood cells, as it is difficult to distinguish cells under the microscope. There should be around 7×10^7 cells ml⁻¹.
- 7** Dilute cells so that there are 3×10^6 cells per 45 µl.
- 8** 3×10^6 cells per block yield ≈ 18 µg DNA (this is suitable for P1 and YAC cloning).
- 9** Dilute 1 : 1 with 1.5% low-melting-point agarose (SeaPlaque), and aliquot 90 µl into precooled block formers.
- 10** When blocks have set, transfer to 0.4 M EDTA, pH 7.5, 1% Sarkosyl, and 2 mg ml⁻¹ proteinase K (up to 25 blocks per 50 ml). Incubate on rocker at 50 °C for 48 h.
- 11** Wash blocks twice in 1 × TE at 50 °C then once in 1 × TE + 40 µg ml⁻¹ PMSF at 50 °C and finally twice in 1 × TE at room temperature. Blocks should be stored in 10 mM Tris (pH 7.5), 50 mM EDTA at 4 °C and washed in 1 × TE thoroughly (e.g. 3 times for 30 min each at room temperature on a rocker) before use.

Protocol 77 High molecular weight liquid DNA preparation

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

(In addition to those for Protocol 76)

- TEN9: 50 mM Tris-HCl (pH 9.0), 100 mM EDTA (pH 8.0–9.0), 200 mM NaCl
- SDS
- phenol
- phenol/chloroform
- sodium acetate
- ethanol

Method

Proceed as for Protocol 76 up to and including step 6.

- 7** Pellet cells after counting and resuspend cell pellet in a small volume of PBSA.
- 8** Add TEN9 (10 ml for every 2×10^7 cells), SDS to a final concentration of 1%, and proteinase K to a final concentration of 0.5 mg ml⁻¹. Make up allowing for addition of 10% SDS to give final concentration of 1% SDS. For example, for a 2.5 ml cell pellet, add:
 - 5 ml 10% SDS;
 - 2.5 ml 10 mg ml⁻¹ proteinase K;
 - 40 ml TEN9.
- 9** Rock at 50 °C overnight.
- 10** Divide solution into 25-ml aliquots in 50-ml Falcon tubes, add 25 ml phenol and rock for 1 h at room temperature.
- 11** Separate phenol and aqueous phases by spinning for 10 min at 3000 r.p.m. in a Beckman J6B centrifuge, and transfer aqueous phase to a fresh tube using a wide-mouthed pipette.
- 12** Extract a second time with phenol (without rocking), once with phenol/chloroform/isoamylalcohol (25 : 24 : 1) (or if necessary, repeat until the solution is clear) and once with chloroform.
- 13** Add sodium acetate (pH 6.0) to 0.3 M and 2 vols of 100% ethanol (or 0.8 vols of 100% isopropanol).
- 14** Loop out DNA with a sterile loop, wash in 70% ethanol by swirling gently. Allow ethanol to evaporate without overdrying the DNA. Allow DNA to resuspend in 1 × TE (pH 8.0) by rocking for at least 24 h at 4 °C.

Protocol 78 YAC library construction

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Overview

- (a) Vector arms preparation: example digest for pYAC4 or pCGS966 vectors
- (b) Production of clonable genomic DNA by partial digests of genomic DNA using *EcoRI*/*EcoRI* methylase
- (c) YAC preparation by ligation
- (d) Size selection of DNA
- (e) Preparation of YACs for transformation
- (f) Preparation of yeast spheroplasts
- (g) Transformation of yeast spheroplasts

Materials

Solutions and media for the spheroplasting and transformation steps (Protocols 78f, 78g) are made on the eve or the day of a transformation from the following stocks:

- 1 M Tris, pH 7.6
- 2 M sorbitol
- YPD (1% yeast extract, 2% peptone and 2% dextrose)
- 1 M sodium citrate, pH 5.8 (0.91 M sodium citrate, 0.093 M citric acid)
- 0.5 M Na-EDTA, pH 8.0
- 1 M CaCl_2

The above are autoclaved and stored at room temperature.

20×AMINO ACID AND ADENINE MIXTURE (FOR PROTOCOL 78g)

- 400 mg l⁻¹ each: adenine, arginine, isoleucine, histidine, lysine, methionine
- 1200 mg l⁻¹ leucine
- 1000 mg l⁻¹ phenylalanine
- 3000 mg l⁻¹ valine
- 600 mg l⁻¹ tyrosine

Dissolve separately in $\frac{1}{20}$ of the final volume, adding NaOH until dissolved.

AS SEPARATE STOCKS (FOR PROTOCOL 78g)

- 8 mg ml⁻¹ tryptophan
- 5 mg ml⁻¹ uracil

The above stocks are autoclaved and stored at 4 °C.

20× YEAST NITROGEN BASE

- 13.6% yeast nitrogen base (without amino acids, Difco) filtered (0.4- μ m filter) and stored at 4°C

(a) Vector arms preparation

The quality of the vector arms preparation is absolutely critical to the overall success of YAC library construction and therefore all reasonable measures should be taken in quality control of the vector DNA. In the protocol described here, dephosphorylated vector arms are ligated to digested genomic DNA in agarose. As in many cases, the amount of genomic DNA is limiting in each experiment and the yield of clones per microgram of ligated DNA is low; a large excess of vector arms DNA is used in every ligation to maximize clonable ligation products and minimize coligation of genomic DNA. Typically, 100 μ g vector arms DNA are used per ligation. To maximize consistency from one cloning experiment to another, it is advisable to prepare large (mg) quantities and to perform quality controls in batch. Vector DNA should be of CsCl gradient purity or equivalent for library construction purposes.

Analytical vector digests. It is always good practice to perform some analytical digests of the vector DNA to confirm that the plasmid is intact. This is particularly relevant to YAC vectors as one of the inverted repeats of the Tetrahymena telomeric sequences is frequently deleted during culturing in *E. coli*. It is often labour saving to perform a quick analytical digest of the vector DNA before CsCl gradient purification. The correct enzymes to use for an analytical digest have to be determined for each cloning vector. The most commonly used vector to date has been pYAC4 and for this vector three digests are recommended: *Eco*RI is the cloning site and *Eco*RI digestion should linearize the plasmid, *Eco*RI/*Bam*HI digestion releases both vector arms and removes the stuffer fragment between the telomeric sequences of the two vector arms, and *Hind*III digestion produces four DNA bands including a small doublet on electrophoresis. When there has been some deletion of one of the telomeric sequences a fifth band is visible under the doublet. In this case the plasmid preparation should not be used for YAC library construction.

The same analytical digests also apply to the pCGS966 vector, which has several advantages over pYAC4 in respect of postlibrary construction clone manipulation [64].

EXAMPLE DIGEST FOR pYAC4 OR pCGS966 VECTORS

Materials

- plasmid vector DNA
- restriction enzymes *Eco*RI and *Bam*HI (New England Biolabs)
- calf intestinal alkaline phosphatase (CIP) (Boehringer Mannheim)
- nitrilotriacetic acid (NTA) (Sigma)

- dextran (T40) (Pharmacia)
- phenol/chloroform/isoamylalcohol (25 : 24 : 1)
- chloroform/isoamylalcohol (24 : 1)
- 3 M sodium acetate pH 5
- isopropanol
- TE (see Protocol 76)

Method

- 1** Digest CsCl-purified plasmid vector at 500 ng μl^{-1} with *EcoRI* and *BamHI* (100 U ml^{-1}) at 37 °C for 3 h. Check for complete digestion.
- 2** Heat denature enzymes at 68 °C for 10 min and re-equilibrate to 37 °C.
- 3** Dephosphorylate 5'-ends of DNA with CIP at 1 U per pmol of DNA 5'-ends at 37 °C for 30 min.
- 4** Denature phosphatase by adding NTA to 15 mM and incubating at 68 °C for 10 min.
- 5** Cool to 37 °C, add dextran (T40) to 100 $\mu\text{g ml}^{-1}$ and extract twice with an equal volume of phenol/chloroform/isoamylalcohol (25 : 24 : 1) and once with an equal volume of chloroform/isoamylalcohol (24 : 1).
- 6** Add sodium acetate to 0.1 M and precipitate DNA with equal volume isopropanol.
- 7** Resuspend DNA in TE at 2 $\mu\text{g ml}^{-1}$ and store at -20 °C.

(b) Production of clonable genomic DNA by partial digests of genomic DNA using *EcoRI*/*EcoRI* methylase

There are several digestion possibilities for the production of clonable genomic DNA. The most favoured system is partial digestion using a competitive reaction with *EcoRI* and *EcoRI* methylase. The size of digest products is most conveniently controlled by varying the amounts of *EcoRI* methylase while keeping *EcoRI* concentration constant.

Materials

- genomic DNA prepared in agarose blocks (see Protocol 76)
- *EcoRI* (New England Biolabs)
- *EcoRI* methylase (New England Biolabs)
- *EcoRI* methylase buffer (New England Biolabs)
- bovine serum albumin (BSA) (Sigma)
- spermidine (Sigma)
- EDTA
- proteinase K (BDH-MERCK)
- phenylmethylsulphonylfluoride (PMSF) (Sigma) (*Caution: extremely toxic*)

Method

- 1** Digest genomic DNA in agarose blocks with *EcoRI* to yield an average fragment size of ≈ 500 kb. To determine the required conditions the blocks and enzymes need to be titrated against one another with a range of *EcoRI/EcoRI* methylase ratios (suggested range: 1 : 40–1 : 320).
Reaction mixture:
 - 90 μ l agarose block;
 - 50 μ l BSA (10 mg ml⁻¹);
 - 50 μ l 10 \times methylase buffer;
 - 13 μ l spermidine (100 mM);
 - 2.5 μ l *EcoRI* (4 U μ l⁻¹);
 - 1–8 μ l *EcoRI* methylase;
 - 290 μ l water.
- 2** Incubate on ice for 1 h and then at 37 °C for 4 h.
- 3** After the digest is completed, the blocks can be loaded directly onto a pulsed-field gel for size selection prior to ligation (for gel conditions see below), or processed according to the following steps.
- 4** Add 55 μ l EDTA (0.5 mM) to stop the reaction and add 62 μ l proteinase K (10 mg ml⁻¹). Incubate at 50 °C for 30 min to digest the enzyme.
- 5** Wash once in 1 \times TE at 50 °C for 30 min on a rocker.
- 6** Incubate in 1 \times TE + 0.04 mg ml⁻¹ PMSF (1 ml per block) at 50 °C for 30 min on a rocker, to inactivate the proteinase K. PMSF is prepared by dissolving to 40 mg ml⁻¹ in isopropanol at 68 °C.
- 7** Wash twice in 1 \times TE at 50 °C for 30 min on a rocker.
- 8** The genomic DNA is now ready for ligation (see Protocol 78c) or size selection prior to ligation (the procedure for size selection before ligation is the same as that carried out after ligation as given in Protocol 78d).

(c) YAC preparation by ligation

Materials

- genomic DNA in agarose blocks
- vector DNA
- 1 \times ligation buffer: 50 mM Tris-HCl (pH 7.6), 30 mM NaCl, 10 mM MgCl₂, 1 \times polyamines (0.75 mM spermidine, 0.30 mM spermine)
- ligation mix: 22 U μ l⁻¹ T4 DNA ligase (New England Biolabs), 5 mM ATP, 10 mM DTT in 1 \times ligation buffer
- polynucleotide kinase (New England Biolabs)

Method

- 1** Wash agarose blocks three times in $1\times$ ligation buffer at room temperature for 30 min.
- 2** Pour off all liquid and place six blocks into a 1.5-ml tube, add an equal amount of vector DNA (i.e. genomic/vector DNA 1 : 1 by mass, i.e. approximately a 500 molar excess of vector arms) and melt agarose blocks at 68 °C for 10 min.
- 3** Allow genomic and vector DNA to equilibrate at 37 °C for 2 h.
- 4** Add $\frac{1}{10}$ volume premade ligation mix.
- 5** Remove two 10- μ l aliquots from each sample, and to one add 1 μ l polynucleotide kinase (10 U μ l⁻¹) as a ligation control. The control is run on a standard 1% agarose gel, and in a successful ligation a ladder of bands will be seen due to the concatamerization of vector arms. The remaining sample is run on a PFG for size selection.
- 6** Ligation is carried out at 20 °C for 16 h.

(d) Size selection of DNA

Size selection is carried out by PFGE in a 1% low-melting-point agarose gel in $0.5\times$ TBE.

Materials

- low-melting-point agarose gel (SeaPlaque, CTG)
- $5\times$ TBE: 54 g l⁻¹ Tris base, 27.5 g l⁻¹ boric acid and 20 ml l⁻¹ 0.5 M EDTA (pH 8.0)
- ethidium bromide
- gel comb (BioRad)

Method

- 1** Tape the teeth of a BioRad gel comb together to form one long trough for each ligation mixture in the gel. Melt the ligation mixture at 68 °C and load it carefully with a cut-off blue Gilson tip into a trough. Either side of the ligation mixture, load one well with ligation mixture and another with yeast chromosome size markers on the outside.
- 2** Seal the wells with molten low-melt agarose to prevent samples escaping.
- 3** Run the gel at 160 V with 30 s switch time, $120\times$ field angle, at 14 °C for 16 h. (These conditions retain fragments > 450 kb in the limiting mobility fraction.)
- 4** After electrophoresis cut the size markers and the single sample wells off the gel using a sterile scalpel and stain in ethidium bromide

solution. Keep the rest of the gel under electrophoresis buffer to prevent drying.

- 5 Visualize the limiting mobility fraction under UV light and cut it out of the single sample lanes.
- 6 Reassemble the gel slices and cut out the limiting mobility fraction of the large sample troughs using a sterile scalpel.
- 7 Stain the rest of the gel in ethidium bromide and take a photograph which serves as record of the fraction of sample DNA in the limiting mobility.

(e) Preparation of YACs for transformation

Materials

- TENP buffer: 10 mM Tris-HCl (pH 7.6), 20 mM EDTA, 30 mM NaCl + polyamines (as in ligation buffer in 3C)
- agarase (Sigma)

Method

- 1 Wash the excised gel slices three times in TENP buffer for 30 min at room temperature.
- 2 Transfer the slices into 1.5-ml tubes and melt at 68 °C for 10 min.
- 3 Allow samples to cool to 37 °C and add agarase (10 U μl^{-1}) to 50 U ml^{-1} and incubate at 37 °C for 3 h.

The DNA is now ready for transformation into yeast spheroplasts. At this point the DNA can be stored at 4 °C for a day or so. For long-term storage the DNA should be kept at 4 °C in solid agarose. After long storage it will be necessary to re-equilibrate the gel slices in TENP buffer before use.

(f) Preparation of yeast spheroplasts

The strain of *S. cerevisiae* used as host in most YAC cloning experiments is AB1380 [14]. It has several auxotrophies, two of which are complemented by the YAC vector. In the case of the pYAC4 vector these are the *trp1* and *ura3* mutations.

Materials

- *S. cerevisiae* AB1380 cells
- YPD medium: 1% yeast extract, 2% bactopectone, 2% dextrose. Make up with 2% agar for plates
- sorbitol
- SCE: 1 M sorbitol, 0.1 M sodium citrate (pH 5.8), 10 mM EDTA (pH 7.5)

- β -mercaptoethanol
- lyticase (Sigma)
- STC: 1 M sorbitol, 10 mM Tris-HCl (pH 7.6), 10 mM CaCl_2
- swinging bucket centrifuge (Beckman J6B)
- haemocytometer

Method

- 1** Streak AB1380 cells out from a frozen culture onto YPD agar plates and incubate at 30 °C for 48 h.
- 2** Use a single colony from the fresh plate to inoculate a 10 ml standing culture in YPD liquid medium and incubate at 30 °C for 24 h.
- 3** Use 250 μl of the standing culture to inoculate a 250 ml YPD liquid culture in a 500-ml conical flask, incubate and shake at 30 °C on an orbital rocker at 50 cycles per min.
- 4** Grow the cells to an OD_{600} of 1.8 (determined on a $\frac{1}{10}$ dilution in YPD) and then pellet in 50-ml aliquots by centrifugation at 3000 r.p.m. for 8 min in a Beckman J6B swinging bucket centrifuge (i.e. 1400 g).
- 5** Resuspend each pellet in 20 ml distilled water and spin again at 3000 r.p.m. for 8 min.
- 6** Resuspend each pellet in 20 ml 1 M sorbitol and spin at 3000 r.p.m. for 8 min.
- 7** Resuspend each pellet in 20 ml SCE and add 46 μl β -mercaptoethanol (14 M). Remove a sample and measure OD_{600} of $\frac{1}{10}$ dilution in water. This serves as a prespheroplasting sample.
- 8** Add 1000 U each of lyticase (Sigma, partially purified) (zymolyase can also be used [36]) to 20 ml of cells and after mixing incubate at 30 °C. Determine the extent of spheroplasting by measuring OD_{600} of a $\frac{1}{10}$ dilution in water. The percentage spheroplasting is taken to be the percentage reduction of the prespheroplast reading of a $\frac{1}{10}$ dilution in water at 600 nm. Spheroplast to 85% (should take \approx 15–20 min) and then spin at 157 g or 1000 r.p.m. in a Beckman J6B swinging bucket centrifuge for 8 min (higher speeds will rupture the cells).
- 9** Gently resuspend the cells in 20 ml sorbitol (1 M) and spin again at 1000 r.p.m. for 8 min.
- 10** Gently resuspend the cells in 20 ml STC. Count the cells under a light microscope using a haemocytometer and calculate the volume to yield a cell concentration of 6.5×10^8 cells per ml.
- 11** Spin cells again at 1000 r.p.m. for 8 min and then resuspended in the calculated volume of STC. Cells are now ready for transformation.

(g) Transformation of yeast spheroplasts

Materials

- spheroplasted yeast cells
- DNA in agarose (as prepared in Protocol 76)
- PEG solution: 20% polyethylene glycol 6000 MW, 10 mM Tris-HCl (pH 7.6), 10 mM CaCl_2
- SOS solution: 1 M sorbitol, 25% YPD, 6.5 mM CaCl_2 , $10\text{ }\mu\text{g ml}^{-1}$ tryptophan, $1\text{ }\mu\text{g ml}^{-1}$ uracil
- YAC regeneration medium (single-selection medium): 1 M sorbitol, 4% dextrose, 0.67% yeast nitrogen base (without amino acids), amino acid supplements (see list at front of protocol), $20\text{ }\mu\text{g ml}^{-1}$ tryptophan, 2% agar
- 15-ml, 50-ml Falcon tubes
- 22×22 cm culture plates

Method

- 1 Add 150 μl of spheroplasted cells to 50 μl of agarosed DNA (≈ 80 ng) in a 15-ml Falcon tube. Allow to sit at 20°C for 10 min.
N.B. DNA must be transformed in 50- μl aliquots as scaling up significantly reduces transformation efficiency.
- 2 Add 1.5 ml filter-sterile PEG solution and allow to sit at 20°C for 10 min.
- 3 Spin tubes at 1000 r.p.m. for 5 min.
- 4 Remove supernatant and resuspend the pellets in 225 μl SOS solution.
- 5 Pool samples into as many 50-ml Falcon tubes as you intend to pour on separate (22×22 cm) plates.
- 6 Incubate at 30°C for 30 min.
- 7 Pour into each tube molten (50°C) YAC regeneration medium and mix cells in by inversion.
- 8 Quickly pour onto 22×22 cm YAC regeneration medium plates that have been prewarmed to 37°C . Tilt the plate to distribute the 'top-agar' evenly before it sets. Note: clones are first grown on this single-selection medium (–uracil only; as the trp promoter on the right arm of the YAC vector is weak and cells often do not grow in double-selection medium (–uracil, –tryptophan) immediately after transformation.
- 9 Allow to set for 10 min at room temperature and then incubate at 30°C for 3–5 days.
- 10 When clones have grown, pick into microtitre plates (see Protocol 79), or use a 40 000 pin replicator device to replicate clones (see Protocol 80).

Protocol 79 Arraying YAC libraries in microtitre plates

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- double-selection growth medium: 4% dextrose, 0.67% yeast nitrogen base (without amino acids), base and amino acid supplements (–uracil, –tryptophan) (see list at front of Protocol 78 for yeast nitrogen base and amino acid mixtures)
- YPD medium (see Protocol 78)
- 4% glycerol
- 96-well microtitre plates
- transfer device (e.g. 12-pinned wheel)

Method

- 1 Pick individual clones into separate 96-well microtitre plate wells, each filled with 100 µl double-selection growth medium per well.
- 2 Incubate at 30 °C for 24–48 h.
- 3 Inoculate all clones into a fresh microtitre plate containing 100 µl YPD medium per well and incubate at 30 °C for a further 24–48 h. The inoculation is done using a transfer device which consists of 96 pins mounted on a plate so that the pin array matches that of the microtitre plate wells. This device is sterilized by immersing in 70% ethanol followed by drying on a hot plate. This step is included because clone regeneration after freezing is more successful when cells are stored in YPD medium.
- 4 Add 100 µl YPD plus 40% glycerol to all wells, mix into grown culture and remove 100 µl into fresh plate as replica. Place on dry ice until solid.
- 5 Store microtitre plates at –70 °C.

A duplicate copy of the microtitre plates should be made, to avoid contamination and decreased viability of original plates during repeated handling.

.....

Protocol 80 YAC library filter lifts

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- YAC clones from Protocol 78 on primary transformation plates
- double-selection medium agar plates: double-selection growth medium: 4% dextrose, 0.67% yeast nitrogen base (without amino acids), amino acid supplements (–uracil, –tryptophan) (see list at front of Protocol 78 for yeast nitrogen base and amino acid mixtures), 2% agar
- SCE (see Protocol 78f)
- Novozym 234 (Novo Biolabs)
- 0.5 M NaOH, 1.5 M NaCl
- 1 M Tris (pH 7.6), 1.5 M NaCl buffer
- 0.1 M Tris (pH 7.6), 0.15 M NaCl buffer
- nylon membrane (Hybond N⁺; Amersham)
- 3MM Whatman filter
- 40 000-pin replicator (available through ICRT, Sardinia House, Sardinia Street, London WC2 A)
- apparatus for UV irradiation

Method

- 1 Dry the primary transformation plates in a flow hood for 5–10 min.
- 2 Sterilize the 40 000-pin replicator in ethanol, drain well and remove remaining ethanol by flaming. Allow the replicator to cool for 2 min.
- 3 Place the replicator onto a solid even surface with the pins facing upwards. Place a primary transformation plate face down onto the pins and apply even pressure over the entire area to ensure that all pins penetrate the agar plate to approximately the same depth. Remove the agar plate evenly from the replicator. If the density of clones on the primary transformation plates is low, it may be desirable to replicate multiple primary library plates onto the pins of the replicator, before transferring the clones onto the fresh agar plate copies. Up to 5000 clones can be replicated onto one plate, still allowing single colonies to be picked after filter lifting. Higher densities may make secondary screening necessary to identify the correct clone.
- 4 Place a dried double-selection plate onto the pins. Applying even pressure, ensure that all the pins have contacted the agar without forcing the pins under the surface. Remove agar plate and incubate at 30 °C for 24–48 h. Up to six replica plates can be made from a single set of inoculated pins.
- 5 After growth, dry replicated plates in flow hood for 5–10 min and then place a dry nylon membrane onto the plate evenly. After 5 min, pull the membrane off the plate by lifting at diagonally opposed corners. Place the membrane colony face up onto a 3MM Whatman filter soaked in SCE + 8 mg ml⁻¹ Novozym 234.

- 6 Incubate the membrane on the 3MM Whatman filter for 1 h at 37 °C to spheroplast the yeast cells.
- 7 Transfer the membrane onto 3MM Whatman filter soaked in 0.5 M NaOH, 1.5 M NaCl and leave to denature at room temperature for 5 min.
- 8 Transfer the membrane onto a 3MM Whatman filter soaked in 1 M Tris (pH 7.6), 1.5 M NaCl and allow to neutralize for 5 min.
- 9 Submerge the membrane in 0.1 M Tris (pH 7.6), 0.15 M NaCl buffer for 2 min. Then air-dry the membrane for 15 min and dry thoroughly between 3MM Whatman filters.
- 10 Cross-link the DNA to the membrane by UV irradiation. The membrane is now ready for prehybridization treatment (see Protocol 90).

.....

Protocol 81 Preparation of YAC agarose blocks

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- *S. cerevisiae* AB1380 clone containing the YAC of interest
- double-selection medium (–uracil, –tryptophan) (see Protocol 78)
- YPD medium (see Protocol 78)
- low-melting-point agarose (SeaPlaque GTG)
- TE50: 10 mM Tris-HCl (pH 7.6), 50 mM EDTA
- Novozym 234 (Novo Biolabs)
- proteinase K (BDH) + sarkosyl, or
- lithium dodecyl sulphate
- SCE (see Protocol 78f)
- DTT
- PMSF (*Caution: extremely toxic*)
- EDTA
- Tris-HCl (pH 8)
- agarose block formers (BioRad)
- centrifuge (Beckman)
- 1.5-ml Eppendorf tubes
- 50-ml Falcon tubes

Method

- 1 Inoculate 10–100 ml of –uracil –tryptophan medium with *S. cerevisiae* AB1380 clone containing the YAC of interest and

incubate at 30°C until the culture is saturated (5×10^7 – 10^8 cells ml⁻¹), about 24–48 h.

- 2 To prepare a frozen aliquot of cells for permanent storage, remove 0.5 ml and mix with 0.5 ml of 30% glycerol in YPD, freeze on dry ice and store at –70 °C. Whenever fresh cells are required, scrape a small amount out of the frozen tube with an inoculating loop and streak onto a YPD plate. Return to –ura –trp selection as soon as possible.
- 3 Prepare molten 1.5% low-melting-point agarose in SCE, cool and maintain at 50 °C. When equilibrated, add Novozym 234–8 mg ml⁻¹ (it is easiest to dissolve the enzyme in SCE before adding to the molten agarose).
- 4 Clean and prepare block formers (see Protocol 76). Place the sealed block formers on a glass plate sitting on ice to facilitate setting of the agarose when added.
- 5 Centrifuge the remaining culture at 3000 r.p.m. (Beckman, J6B) for 10 min at room temperature.
- 6 Discard the supernatant and resuspend the cell pellet in 10 ml of TE50.
- 7 Centrifuge the cells at 3000 r.p.m. for 10 min at room temperature.
- 8 Discard the supernatant and resuspend the pellet in 10 ml SCE.
- 9 Centrifuge the cells at 3000 r.p.m. for 10 min at room temperature.
- 10 Discard the supernatant and resuspend the pellet in SCE at 40 µl ml⁻¹ of original culture (5×10^7 cells). The final yields will be ≈ 1 – 2 µg DNA per block. For more accurate concentrations, count the cells during the previous wash and resuspend to 5×10^7 cells per 40 µl.
- 11 Divide the cell suspension into convenient aliquots (500 µl) in 1.5-ml Eppendorf tubes. Add an equal volume of agarose/Novozym solution to one at a time, rapidly invert to mix, and then pipette 90 µl into each slot of the block formers. Allow to set on ice for 30 min.
- 12 Remove the tape from the block formers and gently push out the blocks using a large inoculating loop, into 50 ml of SCE plus 10 mM DTT (sufficient for up to 100 blocks) in a 50-ml Falcon tube. To allow the cells to spheroplast incubate at 37 °C for 1 h, inverting occasionally.
- 13 To lyse the cells transfer the blocks into 50 ml of either:
 - 0.4 M EDTA (pH 7.5), 1% sarkosyl and 2 mg ml⁻¹ proteinase K; incubate at 50 °C with gentle rocking overnight;
 - or:
 - 1% lithium dodecyl sulphate, 100 mM EDTA, 10 mM Tris-HCl (pH 8); incubate at 37 °C for 1 h; replace with fresh solution and incubate overnight at 37 °C.
- 14 Washing the blocks:
 - *Proteinase K method* The blocks can be stored in this solution at

4 °C indefinitely. Before use the blocks must be washed. Wash the blocks twice in TE50 for 30 min at 50 °C. Then wash with 0.04 mg ml⁻¹ PMSF in TE50, to ensure inactivation of proteinase K (make up PMSF as a 1000×stock in isopropanol). *Caution: PMSF is extremely toxic.* Wash blocks twice more in TE50 for 30 min at room temperature. If the blocks are just to be loaded on a pulsed-field gel then a single rinse in TE50 is sufficient.

- *Lithium method* Wash the blocks four times in TE50 for 30 min at room temperature.

Before the blocks can be used they should be washed (2×30 min) into 1×TE. Blocks are best stored in at least 50 mM EDTA.

.....

Protocol 82

Size fractionation of YACs by pulsed-field gel electrophoresis

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- apparatus for PFGE (e.g. BioRad Clamped Homogeneous Electric Fields (CHEF) system)
- 1% agarose gel (SeaKem GTG, FMC Bioproducts)
- 0.5×TBE (see Protocol 78)
- size markers (e.g. λ-DNA concatemers)
- 0.1 M HCl
- denaturant: 0.5 M NaOH, 1.5 M NaCl
- apparatus and materials for Southern blotting
- 1 M Tris-HCl (pH 7.2), 1.5 M NaCl
- apparatus for UV crosslinking

Method

- 1

Size fractionation is carried out using PGFE on a 1% agarose gel in 0.5×TBE. Yeast chromosomes of the clones act as useful size markers. For further markers however, λ-phage DNA concatamers and yeast YP148 may be useful. YP148 contains pBR322-derived sequence in two of its chromosomes which may be useful when hybridizing Southern blots of YACs. A variety of switch times in the PFGE will yield informative resolution. The simplest is 100 s for 36 h at 4.8 V cm⁻¹ and for more even resolution over a large range of sizes times of 40 s for 16 h followed by 80 s for 12 h and then 110 s for 10 h at 5 V cm⁻¹ will yield a good size separation (90–580 kb).
- 2

We recommend acid depurination of the DNA followed by alkali transfer onto nylon membranes in the form of a Southern blot (see,

e.g. ref. 51 for detailed methods). Submerge the gel in 0.1 M HCl and rock gently for 20 min. Pour off acid and replace with denaturant (0.5 M NaOH, 1.5 M NaCl) and rock gently for a further 20 min. Transfer the DNA by Southern blotting with denaturant to a nylon membrane. After blotting, neutralize the blot in two washes of 1 M Tris-HCl (pH 7.2), 1.5 M NaCl for 5 min, dry and UV crosslink. The blot should be hybridized with the probe used to isolate the YAC. This is to determine the size of the YAC, confirm that it is a single chromosome, and judge the quality of the blocks. The blot should then be probed with labelled total genomic DNA in order to detect any additional YACs (cotransformants) present in the clone.

.....

Protocol 83 Partial restriction digest mapping of YAC DNA

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- YAC block as prepared in Protocol 81
- TE (see Protocol 76)
- TE50 (see Protocol 81)
- restriction enzyme *Bss*HII (New England Biolabs)
- *Bss*HII buffer (New England Biolabs)
- acetylated BSA (Sigma)
- 1.5-ml Eppendorf tubes
- apparatus for PFGE (e.g. BioRad CHEF system)

Method

- 1** Equilibrate one YAC block (prepared as in Protocol 81) in 1 × TE.
- 2** Cut the block into quarters and place each in a separate 1.5-ml Eppendorf tube. The digests are carried out in a final volume of 200 µl which includes the 20-µl volume of the agarose block. Add to each tube 180 µl of buffer composed of 20 µl 10 × *Bss*HII buffer (New England Biolabs or the buffer recommended by your enzyme supplier) and 50 µg acetylated BSA plus sterile distilled water.
- 3** Allow to equilibrate on ice for 30 min.
- 4** Add the enzyme diluted in 1 × restriction buffer to each tube and mix. For *Bss*HII use a range of concentrations 20 U, 0.5 U, 0.15 U and 0.05 U. A range of concentrations should be used in order to detect efficiently all the partial digest products. Allow to equilibrate for 30 min on ice.

- 5 Transfer the tubes to a water bath at the recommended digest temperature (50 °C for *Bss*HI) for 1 h.
- 6 Add 1 ml TE50 and place the tubes on ice.
- 7 Load the digested block onto an agarose gel for PFGE as soon as possible.

The following enzymes will yield partial digests in the above protocol at the concentrations given below, but should be optimized by a titration of enzyme concentrations:

- *M*luI, 0.3 U, 1.0 U, 10 U, and 20 U;
- *N*ruI, *S*alI, *S*acI and *N*otI all at 1 U, 0.1 U and 0.05 U.

The enzyme *S*fiI only gives a partial digestion even at 20 U in the above protocol, cutting very little at lower concentrations.

- 8 Two pulsed-field gels (14×12.7 cm) should be run for accuracy, one to fractionate DNA between 90 and 580 kb and one to fractionate DNA between 10 and 300 kb. Both should be 1% agarose gels in 0.5×TBE and be run at 14 °C. For the former, switching times on the BioRad CHEF system are 40 s for 16 h followed by 80 s for 12 h followed by 110 s for 10 h at 5 V cm⁻¹. For the latter, a linear ramp of switching times from 0.47 s to 26.29 s for 21 h at 6 V cm⁻¹ is used. In addition to λ-ladders and *S. cerevisiae* markers, λ *H*indIII markers should be included on the latter gel.
- 9 After blotting, radiolabelled right and left pBR322 arm probes are hybridized to the filters.

Maps are easily constructed from the sizes of the fragments detected. Other probes can be positioned on the map by probing these blots. In addition to generating fragments by partial digestion of YAC DNA, fragments produced by complete digestion (this can be achieved by digesting with 20 U of enzyme for 6 h or preferably overnight) are useful for positioning probes on the map of the YAC, especially when there is a high density of probes in the region.

.....

Protocol 84 Generation of end-specific probes from YAC clones

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Overview

- (a) Plasmid rescue
- (b) Vectorette probes

(a) Plasmid rescue

Materials

- YAC agarose block
- 1×TE (see Protocol 76)
- restriction enzyme *Xho*I (New England Biolabs)
- restriction enzyme buffer (New England Biolabs)
- T4 DNA ligase (high activity; New England Biolabs)
- ligation buffer (as supplier or see Protocol 78c)
- ATP
- agarase (Sigma)
- phenol/chloroform/isoamylalcohol (25 : 24 : 1)
- chloroform/isoamylalcohol (24 : 1)
- sodium acetate
- ethanol
- *E. coli* XL1-Blue cells (Stratagene)
- ampicillin plates (see ref. 51)
- electroporation equipment (e.g. BioRad Gene Pulser)
- 1.5-ml Eppendorf tubes
- benchtop microfuge

Method

- 1 Equilibrate one 80- μ l YAC agarose block in 1×TE.
- 2 Digest the block with 50 U of *Xho*I in a 200- μ l reaction (including the 80- μ l volume of the block) with the buffer recommended by the enzyme supplier. Digest at 37 °C for 3 h.
- 3 Wash the block in 50 ml of 1×TE for 30 min to 1 h.
- 4 Transfer to a clean 1.5-ml Eppendorf tube and equilibrate with 1 ml 1×T4 DNA ligation buffer (as per enzyme supplier, but without ATP) at 4 °C for 1 h.
- 5 Then add 100 μ l of fresh 1×ligation buffer and melt the block at 68 °C for 15 min.
- 6 Mix and cool to 37 °C and then add 1 μ l 100mM dATP and 400U T4 DNA ligase (i.e. 1 μ l of the high-activity enzyme from New England Biolabs). Incubate at 37 °C for 1h.
- 7 Heat at 68 °C for 15 min and then cool to 37 °C.
- 8 Add 20 U of agarase and incubate at 37 °C for 1 h.
- 9 Extract twice with an equal volume phenol/chloroform/isoamylalcohol (25 : 24 : 1) and once with an equal volume of chloroform/isoamylalcohol (24 : 1).
- 10 Add sodium acetate to 0.3 M and then precipitate DNA with 2 vols 100% ethanol (−20 °C). Spin in a benchtop microfuge for 15 min, remove supernatant, wash once with 500 μ l 70% (v/v) ethanol and air dry for 15 min.

- 11 Resuspend the DNA in 4 µl sterile distilled water.
- 12 Transform 2 µl DNA by electroporation (we use the BioRad Gene Pulser apparatus) into electrocompetent *E. coli* cells such as XL1-Blue. After 1 h incubation in 1 ml growth medium at 37 °C, plate onto plates containing ampicillin (50 µg ml⁻¹) to select recombinant clones.

The probes generated should be mapped back to the YAC and any other panels by hybridization.

(b) Vectorette probes

This method is described in detail by Riley *et al.* [41]. It uses a linker cassette ligatable to a number of sites for frequent-cutting restriction enzymes. A YAC block is digested and ligated with such a linker and vector-specific primers are then used in PCR reactions with a primer specific to the linker. The design of the linker is such that priming from the linker cannot occur until the linker is copied by a vector-primed strand, ensuring that only the vector ends are amplified. The main advantage of this method is that both ends of the YAC can be obtained.

Linker cassettes can be synthesized easily, and when doing this it is necessary to treat the top strand of the vectorette oligonucleotide with polynucleotide kinase and then anneal the two strands [41]. Standard PCR conditions that work well are as follows:

- 94 °C for 5 min;
- 39 cycles of:
 - denaturing at 93 °C for 1 min;
 - annealing at 65 °C for 1 min;
 - polymerization at 72 °C for 3 min;
 - final polymerization at 72 °C for 5 min.

Protocol 85 Cosmid library construction

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Overview

- (a) Preparation of vector DNA
- (b) Preparation of insert DNA
- (c) Ligation
- (d) Host cell preparation
- (e) DNA packaging and transfection
- (f) Plating cosmid libraries

(a) Preparation of vector DNA: example digestion for Lawrist 4**Materials**

- CsCl-purified plasmid vector
- restriction enzyme *ScaI* (New England Biolabs)
- calf intestinal alkaline phosphatase (CIP)
- nitrilotriacetic acid (NTA)
- phenol/chloroform/isoamylalcohol (25 : 24 : 1)
- chloroform/isoamylalcohol (24 : 1)
- 3 M sodium acetate pH5
- isopropanol
- 70% ethanol
- T4 polynucleotide kinase
- TE (see Protocol 76)
- restriction enzyme *Bam*HI (New England Biolabs)
- EDTA
- Dextran T40
- benchtop centrifuge

Method

- 1 Digest CsCl-purified plasmid vector at 500 ng μ l with *ScaI* (100 U ml) for 1 h at 37 °C.
- 2 Heat denature enzyme at 68 °C for 10 min and re-equilibrate to 37 °C.
- 3 Dephosphorylate with CIP (1 U μ mol⁻¹ DNA 5'-ends) at 37 °C for 30 min.
- 4 Inactivate phosphatase by adding NTA to 15 mM and incubating at 68 °C for 10 min.
- 5 Cool to 37 °C and extract twice with an equal volume phenol/chloroform/isoamylalcohol (25 : 24 : 1) and once with an equal volume of chloroform/isoamylalcohol (24 : 1).
- 6 Add 0.1 vol. of sodium acetate and an equal volume isopropanol and spin in a benchtop centrifuge at top speed for 20 min.
- 7 Remove the supernatant and wash the pellet with 500 μ l 70% (v/v) ethanol, air-dry the pellet and resuspend DNA in TE at 1 μ g μ l⁻¹. Assess dephosphorylation by ligation with and without T4 polynucleotide kinase, checking the ligation products on an agarose gel (see, for example, Protocol 78c, step 5).
- 8 If dephosphorylation is complete then digest vector DNA at 500 ng μ l⁻¹ with *Bam*HI (100 U ml⁻¹ at 37 °C for 1 h) and check digestion on a gel.
- 9 If digestion is complete, add EDTA to 15 mM final concentration and heat to 68 °C for 10 min.

- 10** Add 100 µg dextran T40 per millilitre and extract twice with an equal volume phenol/chloroform/isoamylalcohol (25 : 24 : 1) and once with an equal volume of chloroform/isoamylalcohol (24 : 1).
- 11** Add 0.1 vol. of sodium acetate and an equal volume isopropanol and spin in a benchtop centrifuge at top speed for 20 min.
- 12** Remove the supernatant and wash the pellet with 500 µl 70% (v/v) ethanol, air-dry the pellet and resuspend DNA in TE at 0.5 µg µl⁻¹.

(b) Preparation of insert DNA by the time-course method

Additional materials

- genomic DNA
- restriction enzyme *Mbol*
- apparatus for agarose gel electrophoresis or PFGE
- ethidium bromide

Method

- 1** Set up a digest with 1 µg genomic DNA and 0.02 U *Mbol* in a total volume of 30 µl.
- 2** Remove a 5-µl aliquot immediately (i.e. 0 min), add 1 µl 0.5 M EDTA (pH 8.0), heat at 68 °C for 10 min and incubate the remaining reaction at 37 °C.
- 3** Take further 5-µl aliquots at 5, 10, 20, 40 and 80 min. Add 1 µl 0.5 M EDTA (pH 8.0) and heat at 68 °C for 10 min.
- 4** Run all the aliquots out on either a 0.35% (w/v) agarose gel at 0.5 V cm⁻¹, or on a 1% agarose pulsed-field gel at 5 V cm⁻¹ with a switching time of 0.5 s in a CHEF apparatus for 16 h [65]. Choose that time point which gives the most desirable digest product size distribution (i.e. 30–50 kb fragments). When visualizing DNA by ethidium bromide staining, it is important to remember that the fluorescence intensity is proportional to the amount of DNA and that therefore long DNA will stain more strongly than an equimolar amount of shorter DNA.
- 5** Repeat a larger scale digest with 10 µg DNA in 150 µl volume. Remove a 3-µl aliquot at the start and the end of the digest to run on an analytical gel.
- 6** Add 3 U alkaline phosphatase to the remaining reaction and incubate at 37 °C for 30 min.
- 7** Add NTA to 15 mM and incubate at 68 °C for 15 min.
- 8** Extract twice with an equal volume phenol/chloroform/isoamylalcohol (25 : 24 : 1) and once with an equal volume of chloroform/isoamylalcohol (24 : 1).

9 Add sodium acetate to 0.3 M and then precipitate DNA with 2 vols of ethanol (-20°C). Spin in a benchtop microfuge for 15 min, remove supernatant, wash once with 500 μl 70% (v/v) ethanol and air dry for 15 min.

10 Resuspend the DNA in TE to a concentration of 0.5 $\mu\text{g } \mu\text{l}^{-1}$.

(c) Ligation

Additional materials

- cosmid vector DNA as prepared above
- insert DNA as prepared above
- ligation buffer: 10 \times ligation buffer: 400 mM Tris-HCl (pH 7.6), 100 mM MgCl_2 , 1 mM DTT, 5 mM ATP
- T4 DNA ligase
- materials for cosmid packaging and plating (see e and f below)

Method

- 1** Ligate 2 μg cosmid vector to 2.5 μg insert DNA in a 20 μl ligation reaction at 15°C overnight:
 - 4.0 μl vector DNA (0.5 $\mu\text{g } \mu\text{l}^{-1}$);
 - 2.5 μg insert DNA;
 - 2.0 μl ligation buffer (10 \times);
 - 1.0 μl T4 DNA ligase (400 U μl^{-1});
 - H_2O to 20 μl .
- 2** Test 1 μl of ligation by packaging and plating (see e and f below). The remaining reaction can be ligated for a further 3 days at 4°C and then frozen in liquid nitrogen and stored at -70°C .

(d) Host cell preparation

Additional materials

- *E. coli* strain as host
- L-agar plates, supplemented with any required antibiotic for selection
- L-broth
- 10 mM MgSO_4

Method

- 1** Streak out the chosen *E. coli* strain on L-agar plates with any relevant antibiotic, if applicable, and grow at the appropriate temperature overnight.
- 2** Pick a single colony into L-broth (+ any relevant antibiotic) and

incubate in an orbital-shaker-incubator with vigorous aeration (300 r.p.m.) at the appropriate temperature.

- 3 Grow the culture to saturation, then chill in ice water and pellet the cells at 4000 r.p.m. for 10 min.
- 4 Resuspend the cells in half the culture volume of cooled 10 mM MgSO₄. Test the plating efficiency with an aliquot of packaged DNA. The cells are usable for at least 2 weeks when stored at 4 °C.

(e) DNA packaging and transfection

Cosmid cloning benefits from the very high efficiency of *in vitro* packaging, and libraries can be constructed from small quantities of material. We recommend packaging using commercial extracts (Gigapack Gold, Stratagene), following the protocol for packaging and plating supplied. If a commercial packaging extract is not available, then extracts can be prepared from the relevant strains of *E. coli* as described by Frischauf [52].

(f) Plating cosmid libraries

Additional materials

- L-broth

Method

- 1 Determine the titre of the packaged cosmids by plating a serial dilution of the packaging reaction. Add an aliquot of the packaging reaction to 1.5 ml of plating cells and allow to adsorb for 15 min at 37 °C.
- 2 Add 15 ml L-broth and incubate at 37 °C with gentle shaking for 1 h.
- 3 Pellet the cells by centrifugation at 4000 r.p.m. for 10 min, resuspend in 1.5 ml L-broth, plate out evenly onto a 22 × 22 cm agar plate containing the appropriate antibiotic and incubate at 37 °C until the colonies are of sufficient size (the time taken for colonies to reach a certain size is host strain dependent).

.....

Protocol 86 P1 library construction

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Overview

- (a) Preparation of insert DNA
- (b) Preparation of vector DNA
- (c) Production of recombinant DNA
- (d) Packaging

(a) Preparation of insert DNA

Reproducible P1 inserts may be obtained by partially digesting DNA in agarose blocks with *Mbol*, varying enzyme concentrations to find optimal partial digest conditions, initiating reaction with magnesium after extensive equilibration of blocks in buffer without Mg^{2+} containing a suitable dilution of enzyme, and digesting for a fixed time (e.g. 2 h). Initially it is advisable to test a wide range of enzyme dilutions (e.g. on a quarter block), then to select the dilution which appears to work best (in which there is a reduction of DNA in wells and limiting mobility and optimally a highlighted region of fragments just above 100 kb). Test a finer range around this dilution and finally digest several blocks in, for example, three different dilutions, run a quarter block from each reaction on an analytical pulsed-field gel, and select best digests.

PARTIAL DIGESTION OF DNA

Materials

- starting DNA source: genomic DNA prepared in agarose at a concentration of $\approx 18 \mu\text{g}$ per block (see Protocol 76)
- restriction enzyme *Mbol* (Life Science Gibco BRL)
- enzyme storage buffer (for diluting *Mbol*): 50 mM KCl, 10 mM Tris-HCl (pH 7.6), 0.1 mM EDTA, 1 mM DTT, 200 $\mu\text{g ml}^{-1}$ BSA, 50% glycerol
- 1 \times TE (see Protocol 76)
- equilibration buffer: 33 mM Tris-acetate (pH 7.9), 66 mM potassium acetate, 0.5 mM DTT, 1 mM EDTA and 2 mM spermidine
- 100 mM $MgCl_2$
- apparatus for PFGE (e.g. BioRad CHEF II)
- 0.5 \times TBE (see Protocol 78)

Method

- 1 Wash blocks in 1 \times TE at room temperature for 3 \times 30 min on a rocker.
- 2 Equilibrate blocks for 4 h at 4 $^{\circ}\text{C}$ in buffer containing 33 mM Tris-acetate (pH 7.9), 66 mM potassium acetate, 0.5 mM DTT, 1 mM EDTA and 2 mM spermidine (modified from O'Farrell *et al.* [66]) with *Mbol* (diluted in storage buffer) at a concentration according to manufacturer's instructions. (For the BRL enzyme, suggested concentrations are 0.04–0.08 U per block.)

- 3 Initiate the reaction by the addition of MgCl_2 to 10 mM and incubate at 37 °C for 2 h.
- 4 Terminate the digestion by the addition of EDTA to 20 mM.
- 5 Run digests on an analytical pulsed-field gel, for example using a CHEF DRII apparatus (BioRad), 1% agarose gel in 0.5×TBE with pulse times ramped from 3 s to 20 s for 16 h at 180 V. Use λ -concatamer and *S. cerevisiae* chromosome markers (FMC Bioproducts) and undigested DNA for comparison. These conditions resolve DNA up to \approx 250 kb.

PHOSPHATASE TREATMENT OF PARTIAL DIGESTS OF INSERT DNA

In accordance with the cosmid cloning protocols given in ref. 56, partially digested insert DNA is treated with CIP before ligating to vector arms, which are not CIP-treated at the cloning site. (Religated vector arms should not present a problem in the P1 cloning system, which positively selects for clones with inserts.)

Materials

- CIP (Boehringer Mannheim)
- dephosphorylation buffer (Boehringer Mannheim)
- 1×ligation buffer: 50 mM Tris-HCl (pH 7.6), 10 mM MgCl_2 , 30 mM NaCl
- 10×ligation buffer (see Protocol 85c)
- T4 DNA ligase (New England Biolabs)
- T4 polynucleotide kinase (New England Biolabs)
- 1×TE (see Protocol 76)
- proteinase K
- PMSF
- Eppendorf tubes

Method

- 1 Equilibrate blocks containing partially digested DNA in 1×TE for 3×30 min at room temperature.
- 2 Treat blocks with CIP at 0.055 U per microgram of DNA in 1×dephosphorylation buffer, and incubate at 37 °C for 3 h.
Example reaction:
4 blocks = 360 μl
67.5 μl 10× dephosphorylation buffer
4 μl CIP (1 U μl^{-1})
water to 675 μl
- 3 Add EDTA to 20 mM and proteinase K to 1 mg ml^{-1} , and incubate at 50 °C for 30 min to terminate the reaction.
- 4 Wash blocks in 1×TE twice for 30 min at 50 and then twice with PMSF at 40 $\mu\text{g ml}^{-1}$ in 1×TE at 50 °C. Soak blocks in 1×TE once at 50 °C to wash out PMSF. (*Caution: PMSF is extremely toxic.*)

- 5 Estimate efficiency of CIP reaction by religation controls. Use a quarter block per control reaction.
- 6 Wash a three-quarter block in 1×ligation buffer for 3×30 min at room temperature on rocker.
- 7 Cut block into three one-quarter pieces and place each piece in an Eppendorf tube. Melt at 68 °C for 10 min, then equilibrate to 37 °C.
- 8 Prepare reaction premixes as in Table 15.1 (volumes are given in microlitres) and prewarm to 37 °C.
- 9 Add premixes to tubes with DNA, mix gently by stirring with cut tip, incubate tubes at 37 °C for 30 min and then at room temperature overnight.
- 10 Heat religation controls at 68 °C for 10 min, then load on analytical pulsed-field gel using the same conditions as for partial digest DNA.

Expect to see no change between unligated DNA control and ligated DNA. In the presence of polynucleotide kinase however, expect to see an increase in molecular weight, that is the majority of DNA is limiting mobility.

INITIAL DNA SIZE SELECTION BY PFGE

This first size selection by PFGE removes a large proportion of insert fragments that are less than 80 kb. But it is likely that some smaller fragments are still trapped with the larger fragments. The second size-selection gel, after ligation of inserts to vector arms, will remove any remaining smaller fragments.

Materials

- TE50 (see Protocol 81)
- 0.5×TBE (see Protocol 78)
- apparatus for PFGE (e.g. BioRad CHEF)
- λ-concatemers

Table 15.1 Reaction premixes for phosphatase treatment of partial digests of insert DNA.

	Unligated	Religated	Religated + kinase
DNA already molten	22.5	22.5	22.5
25 mM DTT	—	1.2	1.2
25 mM ATP	—	1.2	1.2
10×LB	—	3.0	3.0
Polynucleotide kinase	—	—	1.0
Ligase	—	1.0	1.0
Water	7.5	1.0	—

10×LB = 10×ligation buffer; ligase = T4 DNA ligase (New England Biolabs, 400 U μl⁻¹); polynucleotide kinase = T4 polynucleotide kinase (New England Biolabs, 10 U μl⁻¹).

Method

- 1** Load blocks across a trough of a 1% low-melting-point gel (e.g. SeaPlaque GTG) made with $0.5\times$ TBE. Slice blocks lengthways, and load evenly across the trough; for example, nine lanes of a 30-well CHEF BioRad gel comb taped together to accommodate four partially digested and CIP-treated blocks.
- 2** Electrophorese in, for example, a CHEF DRII apparatus for 16 h using a pulse time of 3 s at 180 V. These parameters compress DNA of >100 kb into the region of limiting mobility. Use λ -concatamers (FMC Bioproducts) as markers.
- 3** Marker lanes only should be stained with ethidium bromide and then realigned with the remainder of the gel, to enable the excision of the limiting mobility DNA which is to be cloned.
- 4** Excise the limiting mobility DNA, stain the remainder of the gel to verify correct excision of material and check the integrity of the DNA.
- 5** Excised DNA can either be stored in TE50, or directly ligated and therefore equilibrated in ligation mix.

(b) Preparation of vector DNA

The P1 positive selection vector (pAd10sacB11) has been described elsewhere [58]. Briefly, this vector (31 kb) contains a *sacB* gene on one side of the cloning site and an *E. coli* promoter upstream. The *sacB* gene encodes an enzyme that converts sucrose to levan which is toxic to the cells. This is the basis of the positive selection for recombinant clones in this system, as religated vector arm clones are able to produce levan in the presence of sucrose, and hence cause cell death. The insertion of DNA into the cloning site prevents the transcription of *sacB*, and hence recombinant clones are viable. In the host strain containing this vector, the *sacB* gene is under the control of a P1 C1 repressor, and hence transcription is blocked. The vector may delete sequences or be subject to rearrangements during culturing. It is therefore advisable to grow up several different cultures at a time and thoroughly check out the vector DNA before using it for cloning.

P1 VECTOR DNA PREPARATION

- *E. coli* strain containing P1 vector
- LB-agar + kanamycin
- LB-agar + kanamycin + sucrose
- LB-broth + kanamycin
- CsCl gradient facilities
- materials for packaging reaction (see d)
- restriction enzyme *SpeI* (New England Biolabs)

Method

- 1 Streak out vector-containing bacteria on LB-agar + kanamycin ($25 \mu\text{g ml}^{-1}$) and inoculate a 25-ml overnight culture with a single colony.
- 2 Inoculate the overnight culture into 800 ml L-broth + kanamycin ($25 \mu\text{g ml}^{-1}$) the following day and grow until saturated (12–14 h).
- 3 Isolate DNA using standard alkaline lysis procedures [51], and purify supercoiled DNA on a CsCl gradient.
- 4 To test purified vector, package $1 \mu\text{g}$ (see d) and plate out $\approx 1 \text{ ng}$ on both kanamycin and kanamycin + sucrose plates. For example, package $1 \mu\text{g}$ vector, take $10 \mu\text{l}$ of packaged material and add $100 \mu\text{l}$ of plating cells, and finally 1 ml of LB-broth. After growth for 45 min at 37°C , plate out $20 \mu\text{l}$ of this onto each plate. Expect to see 1000 times more colonies on the kanamycin plate, than on the kanamycin + sucrose plates, if the vector is not deleted or rearranged in the *sacB* gene region.
- 5 The vector preparation should also be tested by digestion with *SpeI*, which releases a 1.7 kb fragment containing the *sacB* gene.

P1 VECTOR ARM PREPARATION

Materials

- restriction enzyme *ScaI* (New England Biolabs)
- EDTA
- CIP Boehringer Mannheim
- 150 mM NTA
- phenol/chloroform/isoamylalcohol (25 : 24 : 1)
- chloroform/isoamylalcohol (24 : 1)
- sodium acetate
- $1\times\text{TE}$
- materials for ligation and gel electrophoresis (see c)
- restriction enzyme *BamHI* (New England Biolabs)

Method

- 1 Linearize the vector with *ScaI* (e.g. $50\text{-}\mu\text{g}$ aliquots in $300 \mu\text{l}$ total final reaction volume with 100 units enzyme for 4 h at 37°C , add EDTA to 15 mM and then heat to 68°C for 10 min).
- 2 After cooling to room temperature add CIP (1 U per pmol of DNA 5'-ends) and incubate at 37°C for 30 min.
- 3 Terminate the dephosphorylation reaction by the addition of NTA to 15 mM and incubate at 68°C for 10 min.
- 4 Extract twice with an equal volume phenol/chloroform/isoamylalcohol (25 : 24 : 1) and once with an equal volume of chloroform/isoamylalcohol (24 : 1).

- 5** Add sodium acetate to 0.3 M and then precipitate DNA with 2 vols ethanol (-20°C). Spin in a benchtop microfuge for 15 min, remove supernatant, wash once with 500 μl 70% (v/v) ethanol and air-dry for 15 min.
- 6** Resuspend the DNA in $1\times\text{TE}$ to a concentration of $0.5\text{ }\mu\text{g }\mu\text{l}^{-1}$.
- 7** Test the efficiency of dephosphorylation by ligating a small amount of DNA (with and without T4 polynucleotide kinase), and comparing it by gel to unligated DNA.
- 8** Generate vector arms by cleavage with *Bam*HI in the cloning site (e.g. 50 μg of DNA in 300 μl total reaction volume with 100 U of enzyme for 1 h at 37°C).
- 9** Repurify the DNA by phenol/chloroform extractions and ethanol precipitation as above.
- 10** Resuspend the vector arms in $1\times\text{TE}$ at a concentration of 1 mg ml^{-1} . Check by religation controls.

The integrity of the religated material should also be assessed by packaging and infection. Expect to see at least 100 times more colonies on kanamycin plates compared with kanamycin + sucrose plates.

(c) Production of recombinant DNA

LIGATION

The ligation step is performed in agarose and in a similar way to that described for YAC cloning (Protocol 78c).

Materials

- $1\times$ ligation buffer: 50 mM Tris-HCl (pH 7.6), 10 mM MgCl_2 , 30 mM NaCl
- 100 mM ATP
- 100 mM DTT
- 100 mM MgCl_2
- 4 M NaCl
- 1 M Tris-HCl, pH 7.6
- T4 DNA ligase 400 U μl^{-1} (New England Biolabs)

Method

- 1** Equilibrate the DNA slice from the initial size selection gel in four changes of $1\times$ ligation buffer.
- 2** Transfer the DNA slice to an Eppendorf tube and melt at 68°C for 15 min with a eightfold molar excess of vector arms.
- 3** Allow the DNA to equilibrate to 37°C before addition of ligation buffer containing also 1 mM ATP, 1 mM DTT and T4 DNA ligase at 3 U μl^{-1} reaction volume (see example below).

- 4 Stir the mixture gently using a pipette tip cut to 2 mm diameter using a hot sterile blade, and incubate at 37 °C for 1 h, followed by overnight incubation at room temperature.
- 5 Terminate the reaction by the addition of EDTA to 20 mM.

EXAMPLE LIGATION REACTION

- 1300 µl DNA ($\approx 20 \mu\text{g}$)
- 60 µl vector arms ($1 \mu\text{g } \mu\text{l}^{-1}$)
- 14.9 µl 100 mM DTT
- 14.9 µl 100 mM ATP
- 1.9 µl 100 mM MgCl_2
- 1.4 µl 4 M NaCl
- 9.5 µl 1 M Tris-HCl, pH 7.6
- 11.2 µl T4 DNA ligase
- 75 µl water (up to 1490 µl)

SECOND DNA SIZE SELECTION BY PFGE

The DNA is now ready to be size selected again on a second pulsed-field gel. The reproducible size selection of fragments between 80 and 100 kb without trapping smaller fragments has been found to be very dependent on the concentration of the DNA during the size-selection gel. Initially therefore, small amounts of the ligated material (e.g. 50–80 µl) should be electrophoresed in two-lane troughs of a test pulsed-field gel, and the excised DNA run out on an analytical pulsed-field gel to see if the DNA is of the correct size (at least 80 kb). If the DNA shows smaller fragments than expected, a smaller quantity should be loaded. After these conditions have been optimized, a scale-up experiment can be performed to accommodate a larger trough on a pulsed-field gel.

Materials

- low-melting-point agarose (SeaPlaque)
- apparatus for PFGE
- 0.5×TBE (see Protocol 81)
- λ -concatemers

Method

- 1 Melt the ligated DNA at 68 °C for 15 min and load evenly (using a cut tip, see above) into a trough of a 1% low melting point agarose gel, 0.5×TBE. Use λ -concatamer markers as size markers each side of the trough.
- 2 Electrophorese with a 4-s pulse time at 180 V for 16 h (these are slightly different conditions from the initial size-selection gel). The compressed DNA in the limiting mobility should be excised as before.

- 3** Stain with ethidium bromide and realign the remainder of the gel as in the first size selection. Expect to see religated vector arms well below the region of excised DNA.

CONCENTRATION OF SIZE-SELECTED DNA

There are several methods available for DNA concentration and recovery from agarose gels including Qiagen columns (Qiagen), Centricon columns (Amicon), phenol/chloroform extractions and ethanol precipitation, electroelution, *n*-butanol extractions and electrophoresing DNA into a higher concentration agarose. The problems of loss of DNA on membranes and in columns, and shearing of DNA by manipulations as a liquid are often encountered. Using a GELase (Epicentre Technologies), or agarase (Sigma) reaction followed by ethanol precipitation with careful handling is a simple and relatively efficient method, although it appears to be dependent on the purity of the agarose. As batches of low-melting-point agarose can differ greatly, it is best to test each batch of agarose (test melting and resolidification temperatures and check for residues after GELase treatment which interfere with ethanol precipitation).

Materials

- equilibration buffer: 10 mM Tris-HCl (pH 8.0), 20 mM EDTA, 30 mM NaCl
- GELase (Epicentre Technologies) or
- agarase (Sigma)
- ammonium acetate
- ethanol

Method

- 1** Equilibrate the gel slice from the second size-selection gel in the buffer for 3×30 min at room temperature on a rocker.
- 2** Melt equilibrated gel slice at 68 °C for 15 min.
- 3** Cool the DNA to 37 °C and add 0.5 U agarase per 100 µl of molten agarose (or 0.5 U GELase per 300 µl molten agarose) using a cut tip and mix gently.
- 4** Incubate at 37 °C (agarase) or 45 °C (GELase) overnight.
- 5** Check the efficiency of the reaction by cooling the mixture on ice for 30 min and checking for resolidification.
- 6** Using a cut tip, gently mix ammonium acetate (to a final concentration of 2.5 M) into the solution, and incubate on ice for 10 min.
- 7** Spin at 6500 r.p.m. in a benchtop microcentrifuge for 10 min. Transfer the supernatant to a fresh tube using a cut tip, add 3 vols of 100% ethanol and mix in to homogeneity very gently by rolling the tube.

8 Precipitate at -20°C overnight and centrifuge at 14 000 r.p.m. in a microcentrifuge for 30 min at 4°C .

9 Remove the supernatant and wash the pellet with 70% ethanol. After removing as much of the ethanol as possible, add $1\times\text{TE}$ (0.1 original volume) to the DNA pellet, and allow to rehydrate at 4°C overnight, or at 37°C for several hours.

If DNA is to be stored for packaging at a later date, freeze very quickly in liquid nitrogen and store at -70°C .

(d) Packaging

P1 PACKAGING AND RECOVERY OF RECOMBINANT CLONES

P1 packaging extracts can be purchased from DuPont as part of a P1 cloning kit (NEP-1 13). But a large number of packaging extracts may be required for construction of P1 libraries from complex genomes, which is expensive when using the commercial kit. The following methods describe the preparation of head-tail and pacase extracts. These protocols are a modification of those described by Pierce and Sternberg [67].

Packaging extracts are prepared by the heat induction of appropriate P1 lysogens:

- NS3210 (head-tail): $\text{recD hsdM}^+ \text{ hsdR mcrA B}$ (P1 cm c1: 100 rm-am131) (NB: NS3210 is not lysis deficient; it is important to grow it for a limited amount of time after heat induction, or the cells will lyse too soon);
- NS3208 (pacase): $\text{recD hsdM}^+ \text{ hsdR mcrA B}$ (P1 cm-2 c1: 100 rm-am10.1).

Prepare several glycerol stocks of each of these strains, in order to select the best one as assayed by heat induction (compare number of colonies on plates grown at 32°C and 42°C).

P1 HEAD-TAIL EXTRACT PREPARATION

- frozen glycerol stocks of NS3210
- LB-agar containing chloramphenicol ($25\text{ }\mu\text{g ml}^{-1}$)
- LB-broth containing chloramphenicol ($25\text{ }\mu\text{g ml}^{-1}$)
- lysozyme
- 50 mM Tris-HCl, pH 8
- 10% sucrose
- centrifuge (Sorvall)

Method

- 1** Streak out NS3210 from frozen glycerol stocks on LB-agar plates (2 per glycerol stock) containing $25\text{ }\mu\text{g }\mu\text{l}^{-1}$ chloramphenicol. Grow at 32°C and 42°C overnight.
- 2** Choose cells with the highest 32/42 $^{\circ}\text{C}$ ratio. Set up 5 ml culture in LB-broth + $25\text{ }\mu\text{g }\mu\text{l}^{-1}$ chloramphenicol and grow overnight at 32°C .

- 3** Inoculate 2.5 ml of overnight culture each into two prewarmed 250 ml LB-broth + 25 $\mu\text{g } \mu\text{l}^{-1}$ chloramphenicol cultures in 1-litre flasks. Grow at 32 °C to an $\text{OD}_{650} = 0.3$ ($\approx 2\text{--}3$ h). Centrifuge at 4 °C for 10 min at 7000 r.p.m. in a Sorvall centrifuge in a precooled GSA rotor.
- 4** Resuspend cell pellet in 2.5 ml LB-broth and dilute into 250 ml L-broth prewarmed to 42 °C.
- 5** Grow culture for 45 min at 45 °C with vigorous shaking.
- 6** While culture is growing, prepare microfuge tubes for aliquots of the final extract. Make up a 10 mg ml^{-1} lysozyme solution in 50 mM Tris-HCl (pH 8.0), 10% sucrose (filter sterilized, can be stored at -20°C), and aliquot 4 μl into chilled tubes. Prepare a small container of liquid nitrogen for freezing extracts, and a sonicator ready for gentle sonication.
- 7** Remove culture from waterbath after 45 min, place in ice-water bath, begin swirling to cool rapidly and add 62 ml of filtered ultrapure 50% sucrose solution (final concentration $\approx 10\%$). The sucrose helps maintain integrity of cells before centrifugation.
- 8** After 5 min of swirling, cells should be chilled enough to transfer to precooled centrifuge tubes and centrifuge in a cold GSA rotor at 7000 r.p.m., 4 °C for 8 min. Pour off supernatant, and drain excess liquid off pellet for a few minutes (on ice).
- 9** Resuspend each cell pellet in 500 μl of cold 50 mM Tris-HCl (pH 8.0), 10% sucrose, by gentle circular motions mixing in cells with a cut blue Gilson pipette tip (aperture diameter > 2 mm). It is important to act swiftly at this stage to prevent cell lysis. Avoid introducing air bubbles.
- 10** In a Falcon 2059 tube on ice, sonicate suspension gently (e.g. Kontes microultrasonic cell disrupter 2×5 s bursts).
- 11** Using a cut yellow Gilson tip, aliquot 45 μl of extract into chilled Eppendorf tubes containing lysozyme solution. Flick tubes gently and quickly drop into liquid N_2 .
- 12** Store aliquots at -70°C .

P1 PACASE EXTRACT PREPARATION

Additional materials

- frozen glycerol stocks of NS3208

Method

- 1** Streak out NS3208 and grow at 32 °C and 42 °C as for head–tail extract strain.

- 2 Set up a 5 ml overnight culture in LB-broth + 25 $\mu\text{g } \mu\text{l}^{-1}$ chloramphenicol at 32 °C.
- 3 Inoculate two prewarmed 250 ml LB-broth + 25 $\mu\text{g } \mu\text{l}^{-1}$ chloramphenicol cultures each with 2.5 ml of the overnight culture. Grow at 32 °C until cells reach an $\text{OD}_{650} = 0.5$ ($\approx 4\text{--}5$ h). Centrifuge at 4 °C for 10 min at 7000 r.p.m. in a precooled GSA rotor.
- 4 Resuspend cell pellet in 2.5 ml of LB-broth and dilute into 500 ml of LB-broth prewarmed to 42 °C.
- 5 Grow culture at 42 °C for 15 min with shaking at 250 r.p.m.
- 6 Incubate the culture for a further 165 min at 38 °C.
- 7 Chill the culture to 4 °C and pellet the cells at 7000 r.p.m. for 10 min.
- 8 Pour off the supernatant and resuspend the pellet in 1 ml of cold buffer containing 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 50 mM NaCl and 1 mM PMSF. (*Caution: PMSF is extremely toxic.*)
- 9 Sonicate the cell suspension for four 15-s intervals, then centrifuge for 30 min at 17 000 r.p.m. in a Sorvall SS34 rotor.
- 10 Store 20 μl aliquots at -70 °C.

P1 PACKAGING REACTION: TWO-STEP IN VITRO PACKAGING OF P1 CLONES

Two-step *in vitro* packaging is performed as described by Pierce and Sternberg [67] and in the DuPont P1 cloning manual, details of which are also specified here.

Materials

- 10 \times pacase buffer: 100 mM Tris-HCl (pH 8.0), 500 mM NaCl, 100 mM MgCl_2
- head-tail buffer: 6 mM Tris-HCl (pH 8.0), 15 mM ATP, 16 mM MgCl_2 , 60 mM spermidine, 30 mM β -mercaptoethanol, 60 mM putrescine
- DNase-containing phage buffer: 10 mM Tris-HCl (pH 8.0), 10 mM MgCl_2 , 0.1% gelatin, 10 $\mu\text{g } \mu\text{l}^{-1}$ DNase I
- 1 mM dNTPs
- 25 mM DTT
- 50 mM ATP
- pacase extract
- chloroform

Method

- 1 For a 15- μl pacase cleavage reaction, aliquot 9 μl DNA into a fresh Eppendorf tube, and add the following components:
 - 1.5 μl 10 \times pacase buffer;
 - 1.5 μl 1 mM dNTP (each) mix;
 - 1.0 μl 25 mM DTT;
 - 1.0 μl 50 mM ATP.

- 2 Thaw pacase extract, add 1 μ l to each reaction, and mix with cut tip. Replace pacase on dry ice.
- 3 Incubate reaction at 30 °C for 15 min.
- 4 Add to tube 3.0 μ l head–tail buffer and 1.0 μ l 50 mM ATP.
- 5 Transfer (using a cut tip) reaction to a tube containing head–tail extract (this must be thawed immediately before use), and mix by stirring.
- 6 Incubate at 30 °C for 5 min.
- 7 Spin in microcentrifuge briefly to eliminate air bubbles.
- 8 Incubate at 30 °C for a further 15 min.
- 9 Add 120 μ l DNase-containing phage buffer and mix.
- 10 Incubate at 37 °C for 15 min.
- 11 The reaction may now be spun briefly in a microfuge to pellet cell debris, and supernatant transferred to a fresh tube.
- 12 For storage of packaged material, add 10 μ l chloroform. Store at 4 °C.

PREPARATION OF PLATING CELLS

Additional materials

Two *E. coli* strains are available for recovering recombinant DNA after infection with phage lysate: NS3145 [17], and NS3529 [67].

Method

- 1 Inoculate 30 ml of LB-broth with a scraping of frozen glycerol stock of a suitable *E. coli* strain and incubate overnight at 37 °C with shaking at 250 r.p.m.
- 2 Pellet cells at 4 °C and resuspend in 15 ml sterile 10 mM MgSO_4 , 10 mM Tris-HCl (pH 7.6).
- 3 Store at 4 °C (viable for at least 2 weeks).
- 4 Inoculate 0.5 ml cells into 50 ml L-broth.
- 5 Grow at 37 °C shaking at 250 r.p.m. until $\text{OD}_{630} = 0.3$; pellet cells at 4 °C.
- 6 Resuspend cells in 5 ml cold LB-broth + 5 mM CaCl_2 . Store on ice.

P1 ADSORPTION TO HOST CELLS AND INFECTION

Method

- 1 Add 10 μ l packaged material to tube, preferably glass. If chloroform has been added, then evaporate chloroform off at 37 °C for 10 min.

If scaling up in order to plate out a large number of clones for arraying into microtitre dishes (see Protocol 79), it is advisable to plate out clones onto NUNC Bioassay dishes (22×22 cm). Clones can be picked into microtitre dishes containing a freezing medium mixture as previously described in the cosmid section of this chapter (Section 15.7.5) with 25 µg per microlitre of kanamycin, and stored at –70 °C.

- 2 Add 100 µl plating cells, and incubate at 37 °C for 50 min without shaking.
- 3 Add 1 ml LB-broth to the tube, and incubate at 37 °C for 45 min.
- 4 Pellet cells by a brief spin in a microcentrifuge, and resuspend in small volume of LB-broth.
- 5 Plate on LB-agar plate containing 45 µg µl⁻¹ kanamycin and 5% sucrose, grow at 37 °C overnight.

.....

Protocol 87 Extraction of DNA from P1 clones: minipreps

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- LB-agar + kanamycin (25 µg µl⁻¹)
- LB-broth, or
- 2×YT medium (per litre: 16 g yeast extract, 10 g tryptone and 5 g NaCl) or BHI (see ref. 51)
- alkaline lysis I solution: 50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA
- alkaline lysis II solution: 0.2 M sodium hydroxide, 1% SDS
- alkaline lysis III solution: 3 M potassium acetate, 2 M acetic acid
- 1×TE (see Protocol 76)
- ethanol
- 70% ethanol
- equipment for PFGE
- Eppendorf tubes
- isopropanol
- 0.5×TBE (see Protocol 78)

Method

- 1 Streak out P1 clones to single colonies on LB-agar + kanamycin plates.
- 2 Inoculate a single colony into 10 ml of medium (media other than LB-broth can yield a higher cell density, e.g. 2×YT and BHI [51]) and grow overnight at 37 °C.
- 3 Pellet cells (2000 r.p.m. for 10 min in a Beckman centrifuge), and resuspend the pellet in 300 µl alkaline lysis I solution. Transfer resuspended cells to microfuge tubes.
- 4 Add 600 µl alkaline lysis II solution, gently invert tube several times.

- 5 Add 450 µl alkaline lysis III solution, invert tube until mixed well and leave at room temperature until all remaining samples have been completed.
- 6 Centrifuge tubes in a microfuge for 10 min and remove the supernatant to a fresh 2-ml Eppendorf tube. Add isopropanol to the top of the tube. Invert tubes to mix and leave at room temperature for 30 min.
- 7 Centrifuge tubes in a microfuge for 15 min and pour off the supernatant. Allow pellet to dry and resuspend in 1 × TE.
- 8 Add ammonium acetate to 2.5 M final concentration, incubate on ice for 10 min, pellet debris (10 min in a microfuge) and transfer supernatant to a fresh tube. Add 2.5 vols of ethanol to precipitate DNA.
- 9 Wash in 70% ethanol and resuspend the DNA in 20 µl 1 × TE.
- 10 To assess clone size, *NotI*-digested DNA can be run out on a pulsed-field gel using the following conditions (in a Biorad CHEF DR1I): 3–20 s, 20 h, 180 V in 0.5 × TBE (use one-third of DNA preparation).

.....

Protocol 88 Extraction of DNA from P1 clones: maxipreps

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses, see Appendix III.

Materials

- 2 × YT broth
- kanamycin
- alkaline lysis solutions I, II, III as in Protocol 87
- isopropylthio- β -galactoside (IPTG)
- isopropanol
- 1 × TE (see Protocol 76)
- Oakridge tubes

Method

- 1 Inoculate a single P1-infected bacterial colony into 50 ml 2 × YT broth [51] + 25 µg µl⁻¹ kanamycin and grow at 37 °C overnight.
- 2 Inoculate 25 ml of overnight culture into 800 ml 2 × YT culture + 25 µg µl⁻¹ kanamycin (2 per clone).
- 3 Incubate culture at 37 °C for 1 h (or until OD₅₉₀ = 0.1), then induce cells by adding IPTG to a final concentration of 1 mM.
- 4 Incubate culture for a further 4–8 h.

- 5 Harvest cells by transferring to 1-litre centrifuge bottles and spinning in a Beckman centrifuge at 4000 r.p.m. for 20 min at 4 °C. Pour off supernatant carefully.
- 6 Resuspend cells in 10 ml of alkaline lysis I solution, transfer to GSA centrifuge tubes and incubate on ice for 15 min.
- 7 Add 30 ml alkaline lysis II solution while swirling tube very gently. Leave on ice for 5 min.
- 8 Add 22.5 ml of alkaline lysis III solution, shake (for less than 10 s) and leave on ice for 30 min.
- 9 Spin in a Sorvall centrifuge (GSA rotor) for 30 min at 13 000 r.p.m., then transfer the supernatant to a fresh GSA tube (each sample requires two tubes).
- 10 Add 45 ml isopropanol, mix and leave at room temperature for 5 min.
- 11 Spin at 9000 r.p.m. for 15 min, pour off supernatant and let pellet dry for ≈ 20 min.
- 12 Resuspend pellet in 10 ml 1×TE each and combine (therefore 20 ml total per clone) and transfer to Oakridge tubes. Add 19.66 g of CsCl and mix well until all the CsCl has dissolved. Add 1.5 ml of 10 mg ml⁻¹ ethidium bromide.
- 13 Spin in Sorvall centrifuge at 10 000 r.p.m. for 10 min. Transfer supernatant to polyallomer ultracentrifuge quick-seal tubes (25×89 mm) and top up tubes with CsCl solution (49.2 g in 50 ml 1×TE).
- 14 Balance tubes and seal top.
- 15 Spin in ultracentrifuge in VTi 50 rotor at 45 000 r.p.m. for 16 h at 20 °C (set deceleration to zero).
- 16 Harvest bands under long-wave UV light with needle and syringe.
- 17 Extract ethidium bromide with an equal volume of CsCl-saturated isopropanol until all visible colouring is removed and then extract once more.

This protocol can yield ≈ 50 µg DNA but may vary from clone to clone.

Qiagen or Promega Maxiprep columns can also be used for P1 plasmid DNA preparation.

Protocol 89 Competition of probe to remove repetitive sequences

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- radiolabelled probe
- 1 × TE (see Protocol 76)
- genomic DNA (sonicated)
- yeast tRNA
- 1 M sodium phosphate buffer, pH 7.2

Method

- 1 To a radiolabelled probe in 100 µl 1 × TE add 100 µg sonicated genomic DNA of approximate average size 300 bp (100 µg yeast tRNA can be added when hybridizing against YAC DNA) and 1 × TE to a final volume of 176 µl.
- 2 Denature at 100 °C for 5 min.
- 3 Add 24 µl 1 M sodium phosphate buffer (pH 7.2) (final concentration, 0.12 M).
- 4 Incubate at 65 °C for 1–2 h.
- 5 Add competed probe to hybridization buffer.

Protocol 90 Hybridization and washing

(Modified from Church and Gilbert [62].)
For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- prehybridization/hybridization buffer: 0.5 M sodium phosphate (pH 7.2) (0.42 M Na₂HPO₄, 0.16 M NaH₂PO₄), 7% SDS, 1% BSA, 1 mM EDTA, 0.1 mg ml⁻¹ yeast tRNA. The prehybridization buffer can also include 0.5 mg ml⁻¹ sheared human placental DNA
- 40 mM sodium phosphate, 0.1% SDS
- ³⁵S-labelled host or vector DNA

Method

- 1 Prehybridize filters without probe in prehybridization buffer at 65 °C for about 1 h.
- 2 The hybridization buffer is the same as that for prehybridization (except without sheared human placental DNA if added). Add to the hybridization buffer radiolabelled probe to a concentration of ≈ 1 × 10⁶ c.p.m. (Cherenkov) ml⁻¹ (i.e. 0.5 µCi µl⁻¹). Seal hybridization into a plastic bag and incubate at 65 °C for at least 3 h or overnight.

- 3 Washing conditions vary according to the length of the probe and the type of filter used. Generally, YAC filters and short probes (less than 200 bp) are washed less extensively. All washing is performed in 40 mM sodium phosphate, 0.1% SDS. One or two room-temperature washes followed by one or two washes at 65 °C for between 10 and 30 min each are normal. If problems are encountered identifying clones due to lack of background signal, then 1 µCi ³⁵S-labelled host or vector DNA can be added to the hybridization to make the colonies clearly visible.
- 4 Expose to X-ray film for 1–3 days at –70 °C with a single intensifying screen.
-

References

1 Ballabio, A. (1993) The rise and fall of positional cloning? *Nature Genet.* **3**, 277–279.

2 Weissenbach, J., Gyapay, G., Dip, C. *et al.* (1992) A second generation linkage map of the human genome. *Nature* **359**, 794–801.

3 Copeland, N.G., Jenkins, N.A., Gilbert, D.J. *et al.* (1993) A genetic linkage map of the mouse—current applications and future prospects. *Science* **262**, 57–66.

4 Breen, M., Deakin, L., MacDonald, B. *et al.* (1994) Towards high-resolution maps of the mouse and human genomes — a facility for ordering markers to 0.1 cM resolution. *Hum. Mol. Genet.* **3**, 621–627.

5 Breen, M., Deakin, L., MacDonald, B. *et al.* (1994) Towards high-resolution maps of the mouse and human genomes — a facility for ordering markers to 0.1 cM resolution. *Hum. Mol. Genet.* **3**, 621–627.

6 Hudson, T.J., Stein, L.D., Gerety, S.S. *et al.* (1995) An STS-based map of the human genome. *Science* **270**, 1945–1954.

7 Stubbs, L. (1992) Long-range walking techniques in positional cloning strategies. *Mammalian Genome* **3**, 127–142.

8 Strauss, W.M., Dausman, J., Beard, C. *et al.* (1993) Germ line transmission of a yeast artificial chromosome spanning the murine alpha 1 (I) collagen locus. *Science* **259**, 1904–1907.

9 Choi, T.K., Hollenbach, P.W., Pearson, B.E. *et al.* (1993) Transgenic mice containing a human heavy-chain immunoglobulin gene fragment cloned in a yeast artificial chromosome. *Nature Genet.* **4**, 117–123.

10 Schedl, A., Larin, Z., Montoliu, L. *et al.* (1993) A method for the generation of yac transgenic mice by pronuclear microinjection. *Nucleic Acids Res.* **21**, 4783–4787.

11 Maier, E., Hoheisel, J.D., McCarthy, L. *et al.* (1992) Complete coverage of the *Schizosaccharomyces pombe* genome in yeast artificial chromosomes. *Nature Genet.* **1**, 273–277.

12 Hoheisel, J.D., Maier, E., Mott, R. *et al.* (1993) High resolution cosmid and P1 maps spanning the 14Mb genome of the fission yeast *S. pombe*. *Cell* **73**, 109–120.

13 Baxendale, S., MacDonald, M.E., Mott, R. *et al.* (1993) A cosmid contig and high resolution restriction map of the 2 megabase region containing the Huntington’s disease gene. *Nature Genet.* **4**, 181–186.

14 Burke, D.T., Carle, G. & Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806–812.

15 Bellanne-Chantelot, C., Lacroix, B., Ougen, P. *et al.* (1992) Mapping the whole human genome by finger-printing yeast artificial chromosomes. *Cell* **70**, 1059–1068.

16 Schlessinger, D., Little, R.D., Freije, D. *et al.* (1991) Yeast artificial chromosome-based genome mapping: some lessons from Xq24-q28. *Genomics* **11**, 783–793.

17 Sternberg, N. (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proc. Natl Acad. Sci. USA* **87**, 103–107.

18 Pierce, J.C. & Sternberg, N. (1992) A mouse genomic library on the bacteriophage P1 cloning system: organisation and characterization. *Mammalian Genome* **3**, 550–558.

19 Smoller, D.A., Petrov, D. & Hartl, D.L. (1991) Characterization of bacteriophage P1 library containing inserts of *Drosophila* DNA of 7500 kilobase pairs. *Chromosoma* **100**, 487–494.

20 Sanseau, P., Jackson, A., Senger, G. *et al.* (1994) Cloning of the region between HLA-DMB and LMP2 in the human MHC. *Hum. Immun.* **40**, 1–7.

21 Halford, S., Wadey, R., Roberts, C. *et al.* (1993) Isolation of a putative transcriptional regulator from the region of 22q11 deleted in DiGeorge syndrome, Shrintzen syndrome and familial congenital heart disease. *Hum. Mol. Genet.* **2**, 2099–2107.

22 Sternberg, N., Ruether, J. & de Riel, K. (1990) Generation of a 50,000 member human DNA library with an average insert size of 7500 kb in a bacteriophage P1 cloning vector. *New Biologist* **2**, 151–162.

23 Rao, V.B., Thaker, V. & Black, L.W. (1992) A phage T4 in vitro packaging system for cloning long DNA molecules. *Gene* **113**, 25–33.

24 O’Connor, M., Peifer, M. & Bender, W. (1989) Con-

- struction of large DNA fragments in *Escherichia coli*. *Science* **244**, 1307–1312.
- 25 Hosoda, F., Nishimura, S., Uchida, H. & Ohki, M. (1990) An F factor based cloning system for large DNA fragments. *Nucleic Acids Res.* **18**, 3863–3869.
 - 26 Leonardo, E.D. & Sedivy, J.M. (1990) A new vector for cloning large eukaryotic DNA segments in *Escherichia coli*. *Bio/Technology* **8**, 841–844.
 - 27 Shizuya, H., Birren, B., Kim, U.-J. *et al.* (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor based vector. *Proc. Natl Acad. Sci. USA* **89**, 8794–8797.
 - 28 Rackwitz, H.R., Zehetner, G., Murialdo, H. *et al.* (1985) Analysis of cosmids using linearization by phage lambda terminase. *Gene* **40**, 259–266.
 - 29 Kim, U.-J., Shizuya, H., de Jong, P.J. *et al.* (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Res.* **20**, 1083–1085.
 - 30 Ioannou, P.A., Amemiya, C.T., Garnes, J. *et al.* (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* **6**, 84–89.
 - 31 Pavan, W.J., Hieter, P., Sears, D. *et al.* (1991) High-efficiency yeast artificial chromosome fragmentation vectors. *Gene* **106**, 125–127.
 - 32 Srivastava, A.K. & Schlessinger, D. (1991) Vectors for inserting selectable markers in vector arms and human DNA inserts of yeast artificial chromosomes (YACs). *Gene* **103**, 53–59.
 - 33 Smith, D.R., Smyth, A.P., Strauss, W.M. & Moir, D.T. (1993) Incorporation of copy-number control elements into yeast artificial chromosomes by targeted homologous recombination. *Mammalian Genome* **4**, 141–147.
 - 34 Chartier, F.L., Keer, J.T., Sutcliffe, M.J. *et al.* (1993) Construction of a mouse yeast artificial chromosome library in a recombination-deficient strain of yeast. *Nature Genet.* **1**, 132–136.
 - 35 Larin, Z., Monaco, A.P. & Lehrach, H. (1991) Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl Acad. Sci. USA* **88**, 4123–4127.
 - 36 McCormick, M., Shero, J.H., Cheung, M.C. *et al.* (1989) Construction of human chromosome 21-specific yeast artificial chromosomes. *Proc. Natl Acad. Sci. USA* **86**, 9991–9995.
 - 37 Larin, Z., Monaco, A.P., Meier Ewert, S. & Lehrach, H. (1993) Construction of yeast artificial chromosome libraries. *Meth. Enzymol.* **255**, 623–637.
 - 38 Cox, R.D., Meier Ewert, S., Ross, M. *et al.* (1993) Genome mapping and cloning of mutations using yeast artificial chromosomes. *Meth. Enzymol.* **255**, 637–653.
 - 39 Baxendale, S., Bates, G.P., MacDonald, M.E. *et al.* (1991) The direct screening of cosmid libraries with YAC clones. *Nucleic Acids Res.* **19**, 66–151.
 - 40 Kusumi, K., Smith, J.S., Segre, J.A. *et al.* (1993) Construction of a large-insert yeast artificial chromosome library of the mouse genome. *Mammalian Genome* **4**, 391–392.
 - 41 Riley, J., Butler, R., Ogilvie, D. *et al.* (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18**, 2887–2890.
 - 42 Meier Ewert, S., Maier, E., Ahmadi, A. *et al.* (1993) An automated approach to generating expressed sequence catalogues. *Nature* **361**, 375–376.
 - 43 Chumakov, I., Rigault, P., Guillou, S. *et al.* (1992) Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* **359**, 380–387.
 - 44 Feinberg, A.P. & Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 63.
 - 45 Reilly, J.D. & Silva, R.F. (1993) A double-cos-site vector containing a multiple cloning site flanked by T7 and T3 RNA polymerase promoters. *Gene* **124**, 135–136.
 - 46 Ehrlich, E., Craig, A., Poustka, A. *et al.* (1987) A family of cosmid vectors with the multi-copy R6K replication origin. *Gene* **57**, 229–237.
 - 47 Little, P.F.R. & Cross, S.H. (1985) A cosmid vector that facilitates restriction enzyme mapping. *Proc. Natl Acad. Sci. USA* **82**, 3159–3163.
 - 48 Cross, S.H. & Little, P.F.R. (1986) A cosmid vector for systematic chromosome walking. *Gene* **49**, 9–22.
 - 49 Bates, P.F. & Swift, R.A. (1983) Double cos site vectors: simplified cosmid cloning. *Gene* **26**, 137–146.
 - 50 de Jong, P.J., Yokobata, K., Chen, C. *et al.* (1989) Human chromosome-specific partial digest libraries in lambda and cosmid vectors. *Cytogenet. Cell Genet.* **51**, 985.
 - 51 Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (2nd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
 - 52 Frischauf, A.-M. (1991) Construction of genomic libraries in lambda and cosmid vectors. In *Essential Molecular Biology: a Practical Approach* (Brown, T. A., ed.), Vol. 2, pp. 15–38 (IRL Press, Oxford).
 - 53 Hoheisel, J.D., Nizetic, D. & Lehrach, H. (1989) Control of partial digestion combining the enzymes dam methylase and *Mbo*I. *Nucleic Acids Res.* **17**, 9571–9582.
 - 54 Birnboim, H.C. & Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513–1523.
 - 55 Gibson, T.J., Coulson, A.R., Sulston, J.E. & Little, P.F.R. (1987) Lorist2, a cosmid with transcriptional terminators insulating vector genes from interference by promoters within the insert: effect on DNA yield and cloned insert frequency. *Gene* **53**, 275–281.
 - 56 Nizetic, D., Zehetner, G., Monaco, A.P. *et al.* (1991) Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: their potential use as reference libraries. *Proc. Natl Acad. Sci. USA* **88**, 3233–3237.
 - 57 Albertsen, H.M., Le Paslier, D., Abderrahim, H. *et al.* (1989) Improved control of partial DNA restriction enzyme digest in agarose using limiting concentrations of Mg^{2+} . *Nucleic Acids Res.* **17**, 808.
 - 58 Pierce, J.C., Sauer, B. & Sternberg, N. (1992) A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: improved cloning efficacy. *Proc. Natl Acad. Sci. USA* **89**, 2056–2060.
 - 59 Nurminski, D.I. & Hartl, D.L. (1993) Amplification of the ends of DNA fragments cloned into bacteriophage P1. *BioTechniques* **15**, 201–208.
 - 60 Hoheisel, J.D., Lennon, G.G., Zehetner, G. & Lehrach, H. (1991) Use of high coverage reference libraries of

- Drosophila melanogaster* for relational data analysis. A step towards mapping and sequencing of the genome. *J. Mol. Biol.* **220**, 903–914.
- 61 Zehetner, G. & Lehrach, H. (1986) A computer program package for restriction map analysis and manipulation. *Nucleic Acids Res.* **14**, 335–349.
 - 62 Church, G.M. & Gilbert, W. (1984) Genomic sequencing. *Proc. Natl Acad. Sci. USA* **81**, 1991–1995.
 - 63 Lehrach, H., Drmanac, R., Hoheisel, J. *et al.* (1990) Hybridization fingerprinting in genome mapping and sequencing. In *Genome Analysis. Vol. 1: Genetic and Physical Mapping* (Davies, K.E. & Tilghman, S.M., eds), pp/ 39–81 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
 - 64 Smith, D.R., Smyth, A.P. & Moir, D.T. (1990) Amplification of large artificial chromosomes. *Proc. Natl Acad. Sci. USA* **87**, 8242–8246.
 - 65 Charlebois, R.L., Hofman, J.D., Schalkwyk, L.C. *et al.* (1989) Genome mapping in halobacteria. *Can. J. Microbiol.* **35**, 21–29.
 - 66 O'Farrell, P.H., Kutter, E. & Nakanishi, M. (1979) A restriction map of the bacteriophage T4 genome. *Mol. Gen. Genet.* **179**, 421–435.
 - 67 Pierce, J.C. & Sternberg, N. (1992) Using the bacteriophage P1 system to clone high molecular weight (HMW) genomic DNA. *Meth. Enzymol.* **216**, 549–574.

Chapter 16

Long-range physical map construction and the integration of genetic and physical maps

Richard Mott, Andrei Grigoriev* & Hans Lehrach*

Bioinformatics Department, SmithKline Beecham Pharmaceuticals, New Frontiers Science Park, Third Avenue, Harlow, Essex CM19 5AW, UK

*Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin, Germany

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

16.1	Introduction, 422	16.6.1	Ordering probes using distances, 428
16.2	Types of physical mapping data, 422	16.6.2	Ordering probes using heuristics, 429
16.2.1	Single-copy probe hybridization, 423	16.6.3	Fitting clones to a probe order, 430
16.2.2	Multiple-copy probe hybridization, 424	16.7	Detection of chimaeric clones and random noise, 430
16.3	Positional information, 424	16.8	Predictions of experimental progress in genomic mapping by anchoring random clones, 433
16.4	Integration of techniques, 425	16.8.1	Notation and definitions, 434
16.4.1	Inner product mapping, 425	16.8.2	Observations of the progress of mapping experiments, 434
16.4.2	Integration of genetic and physical mapping information, 425	16.8.3	Simplification of theoretical predictions, 436
16.4.2.1	The Généthon human genome map, 425	16.8.3.1	Practical considerations and assumptions, 436
16.4.2.2	The EUCIB mouse-backcross map, 425	16.8.3.2	Number of islands, 436
16.5	Computational requirements for a physical mapping project, 426	16.8.3.3	Length of an island, 436
16.5.1	Experimental strategy, 426	16.8.3.4	Proportion of genome not covered by anchored islands, 437
16.5.2	Fast and easy data entry, 426	16.8.3.5	Number and size of oceans and overlaps, 437
16.5.3	Error checking, 426		References, 438
16.5.4	Feedback, 426		
16.5.5	Data visualization, 427		
16.6	Algorithms for ordering libraries from single-copy probe data, 427		

16.1 Introduction

This chapter first provides an overview of the current state of informatics in physical mapping, and discusses the interesting computational challenges the subject provides. We then focus our discussion on algorithms for building contigs of overlapping clones from single-copy probe hybridization data, together with some mathematical results that predict the progress of a mapping project based on the use of random single-copy probes. Where applicable, the theoretical material is illustrated by its application to mapping the genome of the fission yeast *Schizosaccharomyces pombe* [1,2].

A *physical map* is an ordered sequence of overlapping cloned DNAs that span a genomic region. A *contig* is a contiguous set of spanning clones (i.e. without gaps). A *long-range physical map* covers a significant part of a chromosome with near-continuous clone coverage, in contrast to *short-range physical maps*, which span shorter regions and are usually constructed specifically for positional cloning purposes—for example, for the Huntington's disease gene [3,4].

The basic problem of physical mapping is: given an unordered library of clones from some genomic region, reconstruct an ordered spanning set, such that every part of the region is covered by a clone from the set. Over the next few years accurate physical maps of each of the 23 human chromosomes will be completed, initially at the level of yeast artificial chromosome (YAC) contigs, and then refined down to the cosmid/BAC/PAC level and sequenced [5]. Extensive efforts to map (and in some cases to sequence) other model organisms such as the mouse *Mus musculus*, the nematode *Caenorhabditis elegans*, the puffer fish *Fugu rubripes*, and the yeasts *Saccharomyces cerevisiae* and *Schiz. pombe* are either complete, under way or are planned (see Chapters 26, 29 and 30). Consequently, the contribution of informatics, both in the design and analysis of mapping experiments, will be increasingly important.

The primary computational issues concern the handling and visualization of large, noisy data sets, and how to combine diverse mapping information, such as the integration of genetic and physical maps. The integration of genetic map information is intrinsically important because the initial impetus for physical mapping (at least in humans) has been the cloning of disease genes which are often approximately localized on the genetic map. Other computational issues include designing optimal pooling strategies for clones [6] to reduce the number of experiments required.

All the analyses and maps presented in this chapter were made using the ICRF contig-building package [7]. This is a suite of programs for the display, manipulation and ordering of hybridization data. The package is available by anonymous ftp from ftp.icnet.uk, in the file icrf-public/Genome-Analysis/icrf_contig_v2.tar.Z. The reader should consult the user manual distributed with the package for details on how to run the programs.

16.2 Types of physical mapping data

Physical mapping data fall into two broad categories: *gel digest fingerprinting* and *probe hybridization/sequence-tagged site (STS) content mapping*. The gel fingerprinting approach has been used to construct physical cosmid maps—for example, of *Escherichia coli* [8], *S. cerevisiae* [9] and *C. elegans* [10] (see Chapter 29). Contig-building packages for gel fingerprint data [11,12] compute the likelihood of each pairwise clone overlap under some statistical model for the distribution of restriction sites (typically that the sites occur as a realization of a Poisson process) and then construct contigs for clones with strong overlaps.

Hybridization-based methods have been used to construct maps of *S. pombe* [1,2]. STS content mapping—as used, for example, to construct the YAC overlap map of human chromosome 21 [13], and in the Whitehead Human Genome Map [5]—while differing experimentally from probe hybridization, is equivalent in terms of the raw data it provides. All these approaches are being exploited to construct YAC and cosmid maps, often in combination to provide some independent verification of clone overlaps.

At one level, all hybridization-based approaches give the same information: a series of probes is screened against a clone library and a *hybridization matrix* is created, with the clones represented by the rows and the probes by columns, so that the result of hybridizing a probe to a clone is found in the corresponding row–column intersection of the matrix. Figure 16.1 illustrates some typical hybridization events together with the hybridization matrix they generate, illustrating the confounding effects of repetitive elements and chimaeric clones. For noise-free data, the matrix will be populated with zeros and ones, while for real, noisy data we may think of it as containing grey levels. There is an important distinction between single-copy (SC) probes, such as STSs, and multiple-copy (MC) probes such as short oligonucleotides, in terms of the type of analysis required. SC data provide (ignoring experimental noise) absolute information about clone overlaps,

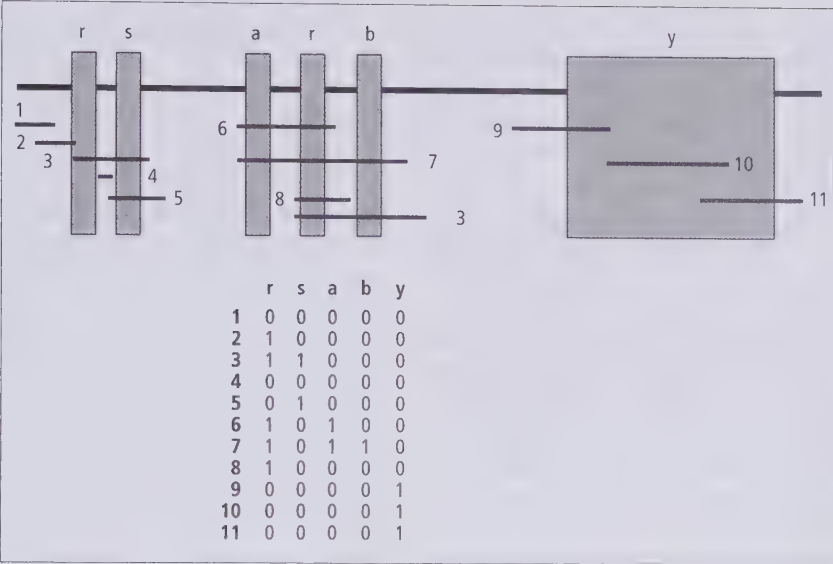


Fig. 16.1 Schematic representation of various hybridization events. The genome is indicated by the thick horizontal line and the clones as short thin lines below, numbered 1–11. Probes are shaded boxes indicating the regions of the genome they span. Below the diagram is the corresponding hybridization matrix. The probes r, s are connected by clone 3 and so would be considered as neighbours. The probe r has a repeat between the probes a, b, causing a possible fork in the

map due to clones 6, 7 and 8. The clone 7 spans a, r, b, so that in this instance it would be possible to remove the probe r without breaking the contig containing a and b. The probe y is a long probe spanning clones 9, 10, 11, showing that clones need not overlap with each other to be detected by the same probe (e.g. 9 and 11). Clone 3 is chimaeric (i.e. contains DNA from two parts of the genome), while clones 1 and 4 are not hit by any probe and so cannot be positioned.

whereas MC data are statistical, like gel fingerprint data, the likelihood that two clones overlap being a function of the number of probes they share. Although the targets of the probes are usually clones covering single genomic segments, they can be multiregion, such as radiation fusion hybrids [14] (see Chapter 14).

In principle, gel fingerprinting can be viewed as a form of MC probe hybridization, where all bands of a given size (or size range) are treated as the same probe. The difference between gel and oligonucleotide fingerprinting (apart from the fact that the former measures sequence length and the latter sequence content differences) is that all information about a single clone is collected in one experiment (the gel fingerprint), with the information about each probe (a fragment of a given size) only emerging when all the clones have been finger-

printed. With oligonucleotide fingerprinting, the converse is true: a single hybridization experiment yields complete information about the probe, but only partial knowledge of a clone's fingerprint. If we think of the data in terms of the hybridization matrix, then an oligo hybridization is a column, and a gel fingerprint a row.

It is useful to classify hybridization data into the categories given in Table 16.1, since their methods of analysis are different.

16.2.1 Single-copy probe hybridization

The simplest case to consider is that of hybridizing SC probes that are much shorter than the target clones. For error-free data, if such a probe hybridizes to two clones, then each clone must overlap with the probe (although not necessarily with each other).

Table 16.1 Categories of hybridization data.

Probe	Example
Short single-copy	STS on cosmid or YAC, cosmid on cosmid
Long single-copy	YAC on cosmid ('Pocket Map' [15])
Short multiple-copy	Oligonucleotide on cDNA
Long multiple-copy	Radiation fusion hybrid Alu-PCR products on YAC

Hence, by hybridizing a sufficient number of SC probes it is possible to link together overlapping clones into contigs and, with a deep library, into a complete continuum of overlapping clones. In Section 16.6 we discuss the theory for evaluating the progress of a mapping project using SC probes, and in Section 16.7 we discuss the algorithms available for ordering noisy data. The theoretical number of probes needed to produce a map will be at least equal to the minimum number of clones needed to span the region of interest, and in practice will be considerably greater, with over half the time probably spent in closing the last few gaps in the contig, as these regions are often hard to clone. A similar situation arises with any other strategy (e.g. gel fingerprinting) predicated on the use of a random library of clones.

Long single-copy probes (e.g. YACs used to hybridize with cosmids) are a rather different case because fewer hybridizations are required to cover the region of interest. The resulting hybridization matrix provides a coarser-grained physical map, since the clones with identical hybridization fingerprints cannot be distinguished, yet may not overlap, but only lie in the same genomic region. The resulting 'pocket map' [15] does, however, contain useful information, since the clones within each pocket can be ordered later with less risk of making wrong connections to clones on other parts of the genome.

16.2.2 Multiple-copy probe hybridization

The great theoretical advantage of MC over SC probes is that the information content of each hybridization is much higher, for the same amount of experimental effort, so that far fewer hybridizations are required [16]. If it is possible to choose statistically uncorrelated probes which each hybridize to a high percentage (ideally 50%) of the target clones, then the number of hybridizations needed to distinguish between N clones is of the order of $\log N$, compared with the order of N with SC probes. However, as the number of expected positives for a probe increases, so does the experimental error rate, and this should be taken into account when designing a mapping strategy (e.g. control clones of known sequence should be included to assay the probe's specificity).

Short MC probes are usually oligonucleotides of 6–20 bp in length, chosen so that they hybridize at a suitable rate with the target clones (the rate being a function of the clone length). These may be used for constructing contigs of overlapping clones, classifying cDNAs [17], etc. They are not considered further in this chapter.

Radiation fusion hybrids (RFHs) can be used either as long MC probes, or as targets for SC probes. Abstractly, a RFH can be thought of as a random collection of large fragments of a chromosome. Radiation hybrid mapping is concerned with ordering SC probes based on their RFH fingerprints. This is very similar to genetic mapping, and many of the same ordering techniques, such as multipoint analysis (see Chapters 3 and 4) and finding orders of probes which minimize the number of obligatory breaks in the RFHs (equivalent to minimizing the number of recombinants), can be applied to the data. [18]. Strictly, RFH mapping (see Chapter 14) does not generate a physical map, since no clone library is screened. However, it may be used as a component in inner product mapping (Section 16.4.1).

16.3 Positional information

Both gel- and hybridization-based methods have the disadvantage that they only give data about a clone relative to its overlapping neighbours, rather than absolute positional information. If there is a non-zero probability of a false positive clone overlap, then any contig constructed purely using overlap information will make a false connection eventually. Consequently, long-range physical map construction is impossible without some form of absolute positional information to check the contigs.

Fluorescence *in situ* hybridization (FISH) is a radically different form of hybridization (see Chapter 9), which provides direct, if imprecise, evidence about the cytogenetic chromosomal location of a clone, and is therefore extremely valuable for anchoring unlocalized clones and the contigs containing them. It can also provide information on chimaerism and repeats if the clone hybridizes to multiple locations. The main disadvantage of FISH (at least when used on metaphase chromosomes) is that it is only accurate to within a chromosome band (say, of the order of 5–10 Mbp) and so it is best used in conjunction with other information (from probe hybridization or gel fingerprinting) to confirm or disprove potential overlaps.

Indirect positional information can be obtained by hybridizing a mapped marker onto a library or by screening with a mapped STS. This is particularly useful if the marker is associated with a disease locus [3]. Similar caveats apply to genetically mapped markers as to FISH—the precision of the genetic map is usually such that local rearrangements of the markers are consistent with the data, so the data should be interpreted carefully.

16.4 Integration of techniques

16.4.1 Inner product mapping

Inner product mapping (IPM) [19] is a potentially powerful combination of SC and MC hybridization, so-called because it is mathematically similar to multiplying together two hybridization matrices. It is a method that propagates mapping information about a few clones onto the unlocalized remainder via a relatively small number of hybridizations of long MC probes. Suppose we have a set of STSs or other SC probes which have known chromosomal locations, and which have been screened against a YAC library. Those YACs which have positive hybridization signals with the probes will consequently be localized too. Now suppose a set of long MC probes (usually Alu-polymerase chain reaction (PCR) products from RFHs [14]; see Chapter 14) are screened against the YAC library. A YAC will be positive with a RFH only if it lies in a chromosomal region covered by the RFH. Each YAC will be characterizable by its RFH fingerprint (i.e. its positive RFHs), and the YACs can be binned into sets with identical fingerprints.

Clearly, those bins that contain positioned YACs will themselves be directly localizable, with a precision related to the numbers of localized YACs and RFH probes (if the RFHs are uncorrelated with a high hit rate then we need of the order of only $\log N$ RFHs to fingerprint N localized YACs uniquely). The surprising part, however, is that the bins that do not contain localized YACs can also have their positions estimated as follows: the RFH-YAC hybridization matrix will imply a YAC fingerprint for each RFH, and so if the YACs are arranged in genome order then the pattern of fragments for each RFH can be inferred from its YAC fingerprint, generating a RFH fragment map. This means that the positions of unlocalized bins of YACs can be estimated by fitting their RFH fingerprints to this map. The actual implementation of this method is more complex than described here because the data tend to be very noisy and so the fingerprints must be interpreted statistically.

The IPM technique was first used to help position unlocalized cosmid contigs in *S. pombe* [7]; in this case a subset of cosmids had been ordered by hybridization to a YAC library, and these cosmids amongst others had been hybridized to the cosmid library, resulting in a set of cosmid contigs. In addition, a minimum-spanning set of YACs had been hybridized to the cosmids, so that each cosmid had a YAC fingerprint. (Thus, the RFHs correspond to YACs and the YACs to cosmids—a sort of

miniaturization of IPM.) Many cosmid contigs could be ordered directly, because they contained mapped cosmids, and some of the other contigs could be positioned approximately via their consensus YAC fingerprints.

16.4.2 Integration of genetic and physical mapping information

The high-resolution genetic maps of the human and other genomes [20,21] are important aids for constructing physical maps. To do this, it is necessary to forge links between markers on the genetic map and clones on the physical map. If the DNA corresponding to the marker is available, then the simplest way to do this is by hybridizing the marker to the clone library, and it is now common for markers to be incorporated into the physical map, and indeed to help to define it. However, this is often only possible for anonymous markers such as STSs, and the positions of important disease-linked loci (see Appendix VII), which may have been positioned by classical genetic mapping of phenotypic traits (see Chapters 1–5) on the physical map must be deduced indirectly by interpolation from their positions between STSs on the genetic map.

16.4.2.1 The Généthron human genome map

The CEPH-Généthron first-generation human genome physical map [22,23] was generated by a combination of techniques: gel fingerprinting of YACs, hybridizations of YAC interAlu-PCR products onto YAC pools, hybridization of mapped markers (from the CEPH genetic map) onto YAC pools, and YAC FISH localizations.

QuickMap, the suite of programs written by Généthron to navigate through this data, is one of the first software packages that attempts to integrate qualitatively different mapping information (see Appendix V for information on how to obtain it). Rather than building long-range contigs, which are likely to be erroneous due to the high frequency of chimaeric YACs in the library (see Section 16.7), the software uses the genetic map as a backbone and suggests tiling paths of YACs connecting pairs of markers or YACs chosen by the user. It treats the information hierarchically: the program attempts to find the best (i.e. shortest legal) connecting path between two objects, preferring marker-to-YAC hybridization, then YAC-to-YAC hybridization, and finally gel fingerprint overlap data.

16.4.2.2 The EUCIB mouse-backcross map

The unification of genetic and physical mapping is even closer in the mouse. The EUCIB mouse

backcross panel [24] (see Chapter 26 for mouse databases) is a set of interspecific hybrid mice, each a backcross between two pure-bred strains, *M. musculus* C57BL/6 (B6) and *M. spretus* (MS). Each animal can be viewed as a genetic chimaera, with alternating sections of B6 and MS (so abstractly each mouse is like a RFH). Many laboratories are using this resource to refine the mouse genetic map, by screening the panel with polymorphic markers.

B1 is a ubiquitous mouse repeat sequence (equivalent to Alu in humans), which can be exploited to make random libraries of B6-specific inter-repeat sequence (IRS) PCR products [25]. These are probed against the EUCIB panel by hybridizing to pools of individual mouse IRS-PCR products. Consequently, if a B6-specific IRS-PCR probe (i.e. with a different sequence in MS) is hybridized to the mouse pools then it should only be positive with those mice that are B6 for the relevant section of the mouse genome. The data obtained from series of screenings with these probes is logically identical to the data observed for binary traits linked to the corresponding loci. Consequently, algorithms for ordering this type of genetic data, or for ordering SC probe hybridization data, are applicable to the ordering of the B6 probes. Once the probes have been ordered, they are hybridized onto B6 mouse IRS-PCR YAC libraries to anchor unlocalized YACs and YAC contigs [26].

16.5 Computational requirements for a physical mapping project

We now turn to some detailed considerations that have been glossed over in the preceding sections. Apart from programs that actually help build the physical map, there are many other aspects to the informatics required by a physical mapping project. Many of these are quite simple, but still make an important contribution. It is essential that there are close links between the informatics and experimental sides of a project, so that the programmers can quickly respond to the needs of the experimenters, and, conversely, so that the experiments can be modified or directed by the results of earlier rounds of hybridizations.

16.5.1 Experimental strategy

The strategy adopted (e.g. random probing, directed contig walking, etc.) must be capable of verification. With random anchoring, the progress can be measured in terms of the growth in the number of contigs, so that the optimal time to switch to a more directed strategy can be determined (see Section

16.8). With walking, it should be possible to verify that each new hybridization is positive with respect to the preceding member of the contig.

16.5.2 Fast and easy data entry

An important bottleneck can be the scoring of the positives on a hybridization from an autoradiogram or phosphorimage and the transfer of the coordinates into a database holding the hybridization data. Although in principle an automated image analysis system is desirable, in practice manual methods are often faster and more reliable if there are only a few positive signals on an image or if there are large variations in signal intensity across an image. We have developed a range of data entry methods, ranging from the fully automated (for oligonucleotide fingerprinting of cDNAs) to semi manual (e.g. for YAC-to-YAC hybridization).

A very useful tool for scoring poor-quality autoradiograms (where there is little background and the overall grid is hard to determine) is to overlay a transparent acetate on which has been printed an idealized grid the same size and configuration as the pattern spotted on the filter. By aligning the acetate to those parts of the image which contain visible background, it is possible to identify the positions of the positives, and with the aid of labelled axes along the sides of the grid, to read off the corresponding microtitre plate coordinates.

16.5.3 Error checking

Well contaminant events are relatively easy to identify, since the probability that two clones from neighbouring wells in a microtitre plate overlap is very low (but not impossible), so that if the positive clones for a given probe include well neighbours then they should be flagged and possibly removed from the analysis. Other forms of sanity checking include the removal of clones or probes with unrealistically large numbers of positives, and, in projects where clones are hybridized as probes, checking if a probe is positive with itself.

16.5.4 Feedback

With rare exceptions, it is not possible (or indeed efficient) to order a data set blind, waiting until all the experiments have been carried out before analysing the results. Preliminary analysis of data can help pinpoint problems with the experiments, potentially saving time and money. The importance of feedback between experimental results and

design of the next round of experiments is dependent on the mapping strategy—a scheme of sampling without replacement or of contig walking is impossible without feedback, whereas a completely random approach is independent.

16.5.5 Data visualization

One of the most important requirements is a method for visualizing the data. The *probe-clone incidence matrix*, with the probes and clones arranged in genome order, is perhaps the ideal way of displaying a synthesis of the raw data and its interpretation because the goodness of fit of the data to the order can be assessed in terms of the strength of the positives aligned along the main diagonal vs. the off-diagonal signals. Such a display tool is invaluable in evaluating the output of different ordering programs, and also for investigating the cause off-diagonal noise, which may not be noise at all, but due to a repetitive probe, a chimaeric clone, or a well contaminant. Other features, such as probes or clones with unrealistically high numbers of positives, are also easy to spot. These effects can be identified by running a program designed specifically to detect them, or by ordering the raw data blind and displaying the (probably partially erroneous) resulting contigs.

As part of the ICRF contig-building package [7], we implemented this display both as a hard-copy PostScript generator, **show**, and as an X-windows application, **xvshow**. All the hybridization matrix figures in this chapter were generated by **show**.

Other forms of display are appropriate once the map has begun to take shape and reasonable estimates of length of clones and their overlaps are possible. Then a display like that in the *C. elegans* database ACeDB [27], with the clones represented as overlapping intervals, is useful. Hybridization positives for each clone or probe can be shown by clicking with a mouse. QuickMap displays contigs graphically as objects linked together by lines or arrows depending on the type of link (e.g. hybridization positive or gel fingerprint overlap).

16.6 Algorithms for ordering libraries from single-copy probe data

For the remainder of this chapter we concentrate on the problem of ordering SC probe hybridization data. Algorithms for ordering MC probe data are described in refs 28–30. For perfect SC data, it is possible to rearrange the order of probes and clones so that the hybridization matrix consists of a

diagonal band of ones, with zeros elsewhere. This is sometimes called the ‘continued ones’ property. There exists a fast algorithm based on interval graphs which will find such an order in time proportional to the number of probes, provided such an order exists [31]. However, the presence of noise almost always means that in practice this is very difficult, so other more robust methods must be used.

Given the inherent problems with clone libraries (chimaerism, repeats, well contaminants, low redundancy, etc.) it is unlikely that map construction will be completely automated in the near future. Building a physical map is an iterative process, with the ordering algorithms suggesting contigs which are manually checked and refined, taking into account other sources of information.

The problem of incorporating positional information from FISH, the genetic map, and so on is still not completely solved. This information may not be consistent, or may be in different coordinate spaces (e.g. FISH data will refer to cytogenetic location, whereas genetic map data will be in a centiMorgan coordinate space). It is not clear what is the best way to impose soft positional constraints on overlap data. An approach to physical mapping using Constraint-Logic Programming in the case of hard positional constraints imposed by the genetic map is described in ref. 32. The approach we adopt, and which works quite well, is to treat all positional information as secondary to the overlap data. The data are first ordered into contigs just using the hybridization data. Then the contigs are ordered and orientated using any positional information attached to any probe or clone in a contig. Inconsistent mapping data (e.g. two probes from different chromosomal regions lying in the same contig) are flagged. It is then up to the experimenter to decide how to resolve such contradictions. Until we know how to compute numerical values for the relative uncertainties of different data (e.g. in the weight given to signals on different autoradiograms, or the accuracy of a FISH result). The resolution of contradictions will require a high degree of human skill to interpret.

We describe two algorithms that are robust enough to order libraries using noisy SC probe hybridization data. One is based on distance measures and simulated annealing, the other on the application of heuristic rules to clean the data into a consistent set.

Both methods order probes rather than clones. Since the number of probes is usually much smaller than the library size, it is more efficient to order the probes first and then fit the clones to the probe order



Fig. 16.2 The YAC map of *S. pombe*. The probes correspond to columns and the clones to rows. All the data were used by the simulated annealing algorithm **probeorder**, whilst the heuristics algorithm **barr** filtered out those probes without a vertical black line (as potentially repetitive), and the grey clones (as potentially coligated). The figure was produced by the **show** program.

automatically. Also, when comparing two probes, we are averaging information over the large number of clones, whereas when comparing two clones, we are averaging over the small number of probes, so a probe–probe comparison has a higher information content than a clone–clone comparison.

A consequence of probe ordering is that a contig is defined as a sequence of probes, rather than clones, and a map as a sequence of probe contigs. This makes for a very compact representation of a map; for example, the *S. pombe* YAC map in Fig. 16.2 can be completely specified by a file listing the probes in order, with the chromosome/contig breaks marked. The probe map can easily be edited manually with a

text editor to modify the positions of misplaced probes using additional information.

16.6.1 Ordering probes using distances

Suppose that two probes, *a* and *b*, have *r* positive clones in common, out of a total of *n* hit by either probe. We define the distance between the probes as $d = (n-r)/n$. This measure is 0 if the probes have identical hybridization patterns, and 1 if they have no positives in common. It is ‘short-sighted’, in the sense that if the probes are more than one clone-length apart, then the estimated distance is always unity. The distance is formally similar to that used in

ref. 18 to estimate the breakage frequency of markers hybridized with radiation hybrids (see Chapter 14), and as an estimate of the meiotic recombinant frequency in genetic mapping (see Chapter 1).

The task of ordering the probes can be cast as 'the travelling salesman problem': find that circular ordering of the probes with the minimum total path length, defined as the sum of interprobe distances. We use simulated annealing [33,34] to find the order of probes with close-to-minimum path length. Since each interprobe distance need be calculated only once, the execution time is dependent primarily on the number of probes, not the number of clones.

The output probe order of the annealing is then broken into a set of probe contigs, with either no or very few clones connecting the last probe of one contig to the first probe of the next. An adjustable cut-off distance value is used to determine where the contig breaks occur. Any probes that have been previously mapped provide a means for ordering and orienting these contigs into their correct positions on the genome, for if two contigs contain neighbouring mapped probes then it is likely that the contigs are adjacent, even if there are no hybridizations linking them. To order the cosmid library of *S. pombe*, the map established by previously ordering the YAC library was used in this way.

Once the order of probes and contigs is established, the clones are fitted to the probe order. All potentially inconsistent hybridizations—that is, any hybridization linking a clone in one contig to a probe in another—are listed. If, after being ordered using a map, a pair of linked contigs is adjacent, then these links are more likely to be genuine. If a clone has also been used as a probe, then the program also checks if the probe and clone are assigned to the same contig and if the probe hits itself.

The algorithm has been implemented in a program called **probeorder** (available in the ICRF contig-building package). It can be used to order any set of single-copy probes that are all approximately the same size, such as cosmid and marker probes on YAC filters, YAC probes on cosmid filters or cosmid probes on cosmid filters. In the case of ordering YAC probes, a modified distance measure is used in which the influence of each cosmid clone is weighted in proportion to $1/n$, where n is the number of YAC probes positive for that cosmid clone. This downweights the effects of clones containing repeats, and which are positive for many YAC probes. The ordering of YAC probes was harder than for other probe types in that the highly variable length of the YACs meant that some YACs were contained entirely within others, and some

YACs were chimaeric, requiring some manual adjustment of the order.

16.6.2 Ordering probes using heuristics

In the case of noise-free experimental data, various simple algorithms, exploiting graph structures or tree-search techniques, can successfully order the library. A general outline for any such algorithm will be as follows:

- 1 for each probe find all neighbouring probes; that is, probes linked by jointly positive clones;
- 2 order all probes relative to their neighbours according to the following procedure (gotos can be replaced by recursion):

```
while an unordered probe exists
  start at some random unordered probe X
  elongation:
    mark current probe as ordered
    find its least/most distant neighbour Y in one direction
    if no unordered neighbours exist
      if only one direction is searched through
        take X as current probe again and change direction
        goto elongation
      else continue (next iteration of while loop)
    if the most distant neighbour Y is found
      mark all probes common for both neighbourhoods as ordered
      (being between these two)
      take this neighbour Y as current probe
      goto elongation
```

In the neighbourhood of a given probe, X , the *most distant neighbour* can be defined either as that probe whose own neighbourhood shares the smallest number of probes with X , and/or as that probe with the smallest number of clones connecting it with X . One can define the *least distant neighbour* of the probe X analogously.

This algorithm will produce a relative order of probes for each contig. The choice between searching for the least or most distant neighbours may depend on the experimental (or even presentational) needs because in the former case the algorithm finds a more detailed and possibly redundant order of probes, while in the latter case it finds a minimal set of probes connected by clones spanning large regions of the genome. Combining both cases may also be useful for checking the consistency of both orders by superimposing and comparing them.

However, it is easily seen that a single false

positive will create a fork in a map. Realistic experimental noise together with the ‘natural’ forks caused by repetitive sequences will result in unpredictable and far from real maps. So a stage of initial filtering of the data becomes a prerequisite.

If a fork results from repeats in the genome (and is therefore likely to violate the neighbourhood rules given below), it is reasonable simply to make a break in the contig because it represents a limit of the experimental technique used and other approaches using longer probes or clones may help to close the gap. In many cases, however, neglecting the data from a probe containing a repeat allows one to find a correct connection to elongate the contig (Fig. 16.1).

In contrast, random false hybridization signals and non-contiguous clones yield additional false neighbours which can be identified by lower numbers of links with a given probe. Neglecting clones producing these links is the simple way of resolving the corresponding forks, although a more careful analysis of hybridization data for such probe pairs and clones linking them could help to reduce the level of noise (see Section 16.7).

We combined with the ordering algorithm a method for finding probes and clones that may cause map forking. A few simple heuristic rules are used to identify suspect clones and probes and to find each probe’s neighbours simultaneously. Then the ‘suspects’ are presented to the user who decides upon removing them from the analysis. After the user’s decision, contigs are built according to the procedure above. The rules for filtering are as follows:

- 1 considering the clones hit by, at most, N probes, the number of neighbours for any probe must not exceed $2(N-1)$;
- 2 the number of neighbours for any probe in any one direction cannot exceed $N-1$;
- 3 for two probes to be neighbours, the number of clones, n , positive for both of them must be >1 (the high coverage of the *S. pombe* library allowed us to use $n=3$).

Thus, the process of ordering a library consists of several iterative stages of filtering out clones that connected pairs of probes less than n times. At each stage, only clones hit less than N times are analysed. If, after filtering the clones, a probe having more than $2(N-1)$ neighbours is found, it is reported as a suspect one and the user may remove it from the analysis and repeat the procedure. During ordering of the probes, a constraint on the number of neighbours in any one direction is checked.

The algorithm has been implemented in two versions (using least and most distant neighbours). The least-distant-neighbour version, **costig**, was

more applicable to ordering the cosmid library under the scheme of sampling without replacement and has a menu-driven interface allowing, among other options, the output of any single contig specified by a probe belonging to it. The most-distant-neighbour version, **barr**, was used for ordering SC probes hybridized to a YAC library.

16.6.3 Fitting clones to a probe order

Once the correct order of probes has been established, it is easy to fit the clones to this probe order, using an algorithm which essentially places each clone on that section of the probe order where it has the highest density of positives. An ordering of N probes imposes a natural integer-valued coordinate system on the genome, in which each probe occupies one of the positions $i=1, 2, 3 \dots N$. It is then sufficient to determine each clone’s start and end coordinates—say, (s,e) . Define a match $a>0$ and a mismatch $b<0$ cost to score the fit $f(s,e)$ of a given clone to the interval (s,e) ,

$$\begin{aligned} f(s,e) &= 0 \text{ if } e < s \\ &= \max(0, f(s,e-1) + ah(e) + b(1-h(e))) \quad (16.0) \\ &\text{otherwise} \end{aligned}$$

where $h(e)$ is defined to be 1 if the clone is positive with the probe at position e and 0 otherwise. The values of (s,e) which maximize $f(s,e)$ define the best-fitting range for the clone.

Given an input order of probes, obtained either manually or by running an ordering program, it is very informative to fit the clones to the probe order and then display the results. Events such as repeats, chimaeras, and so on are often identifiable from the graphic, and the user can make deductions about which clones or probes to exclude from further analysis, and what experiments would help confirm the map. We implemented a program called **reorder**, which, used together with **show**, enables the user to do this. Case Study 16.1 describes the ordering of the YAC map of the *S. pombe* genome.

16.7 Detection of chimaeric clones and random noise

Although they give sufficient accuracy for mapping genomes such as *S. pombe*, where the YAC library contained 47 genome equivalents with only 13% coligation frequency, the algorithms described in Section 16.6 may encounter problems with libraries of higher rates of chimaerism and lower redundancies. For example, the Génethon human genome map [13] is built from a YAC library with about

Ordering the YAC map of *S. pombe*

Although the algorithms described in this chapter are quite general, to set the context we give a brief description of the *Schiz. pombe* libraries and mapping strategy. The haploid genome is 14 Mbp, divided into three chromosomes [48]. Probes were hybridized to cosmid, YAC and P1 libraries. The YAC library comprised 1248 clones with an average insert size of 535 kbp, yielding a coverage of 47 genome equivalents. For the P1 library there were 4056 clones with average insert size 70 kbp—that is, 20 genome equivalents. The cosmid library had an average insert size of 37 kbp. The total cosmid library contained about 8500 clones, a coverage of 23, but most hybridizations were done on a sublibrary of 3000 clones, that is, a coverage of 8.5. Apart from a large tandem rDNA repeat region on chromosome III, the other repeats are confined to the three centromeres.

The main types of probe used on each library were:

- YAC library: whole cosmids, genetically mapped markers, YAC-end probes
- cosmid library: whole cosmids, whole YACs, genetically mapped markers, YAC end-probes, P1 end-probes
- P1 library: whole cosmids, whole P1s, whole YACs, genetically mapped markers.

The basic strategy was top-down: first, order the YAC library by hybridizing cosmids and genetic markers, and then order the cosmid library, using the cosmids hybridized to the YACs as a high-density probe-tagged site (PTS) map. The other cosmid probes were picked by sampling without replacement. The P1 library was ordered in parallel and used to bridge any gaps between the cosmid contigs.

The two algorithms described above were used to order the data. The contigs of the cosmid and YAC libraries found by the two methods were essentially identical except for minor differences in ordering neighbouring probes having very similar or identical hybridization patterns and which could be easily swapped. That was a convincing check for consistency of the resulting map. The map was also verified experimentally by digesting a spanning subset of 40 YACs with *NotI*, and comparing fragment digests with the order of YACs inferred from the hybridization data

Because simulated annealing is a stochastic algorithm, different runs of the program **probeorder** will not necessarily produce the same output probe order. However, we found in practice that in different runs very similar contigs were generated, the differences between runs being confined to contig breaks and to regions in the probe order where the probes were repetitive. If there was no map information available then the order and orientation of the contigs was random, but the order of probes within each contig was stable.

In 10 runs on the complete YAC dataset, using a distance cutoff of $d=0.85$ to determine contig breaks, the 'correct' probe order for both chromosomes I and II was found on

two occasions, each time with the same path length (i.e. sum of interprobe distances) of 7381. The rDNA repeat at the end of chromosome III was always placed incorrectly next to the rDNA repeat at the beginning. On the other runs the final path length was slightly higher, with the maximum value over all runs of 7392, and with errors in the probe orders being confined to the centromeres of one or both of chromosomes I and II (e.g. one half of chromosome II would be joined to chromosome III). The remainder of the probe order (between the each telomere and the corresponding centromere) was correct in all runs. It is straightforward to edit the errors manually.

In this data set the lowest path length found did correspond to the correct order for chromosomes I and II. However, the fact that the other local minima found by the algorithm were all less than 0.15% greater than this value indicates that the path length would not necessarily have a minimum coinciding with the best probe order in other data sets containing many repeats, a fact illustrated by chromosome III. Consequently, it is always worth considering other local minima found by the annealing process, as these may correspond to better alternative probe orderings. Also, other measures of pairwise probe distance/similarity may prove to be better-suited to future applications than the distance used here.

To demonstrate the robustness of the heuristic filtering procedure, the YAC map was constructed using **barr** on the complete YAC data set. During the iterative runs of the program when the parameter *N* (the number of hits per clone) was changed from 2 to 7 and all the 'suspect' probes reported by the program were deleted together, not taking into account the information about repetitive probes. Thus human intervention was excluded in this blind approach, to demonstrate possible problems in cases where there was no genetic or PTS map.

Figure 16.2 shows the result superimposed onto the final YAC map. Clones deleted in the analysis are indicated as grey horizontal lines, with the nondeleted clones shown black. Similarly, the black vertical lines correspond to the ordered subset of probes, while the lines are omitted for the probes that were filtered out, resulting in white-space breaks.

It is clearly seen that all the probes producing blocks of extra positives outside the main diagonal (like those hitting the centromeric regions of all three chromosomes) are successfully filtered out and most (about 80%) of the coligated clones or those containing repeats are excluded from the analysis. The remaining subset of the initial raw data allows the program to reconstruct the complete maps of the chromosomes I and II as well as most of the chromosome III map.

The wide gap in the chromosome III map (accounting for 1 Mbp of the rDNA repeat) is in fact not a gap but an undetected overlap of two groups of clones having five

deleted probes in common. This is an artefact of the 'blind filtering' approach, when information about genetically mapped probes in this region was intentionally ignored and all 'suspect' probes were deleted together. Obviously, if probes in this region were deleted one by one, starting from the probes known as repetitive (containing rDNA or belonging to telomeres in this case), fewer would be deleted and the overlap would be successfully detected.

This example also demonstrates the basic principle of handling the repetitive elements: probes in the region containing this element are deleted one by one until two non-repetitive probes on both sides of the region are found which are connected by a minimum number of clones, linking these probes into a contig. Thus all the probes producing vertical 'blocks' of positives outside the main diagonal in Fig. 16.2 are removed from the analysis and the resulting contigs can be found successfully. In the case when a clone length is less than that of a repetitive element (or a total length of a group of adjacent repeats), a break in the contig is inevitable and can only be closed by using longer clones which can bridge the repeat. Therefore, for highly repetitive genomes it is likely that only maps containing clones from libraries of different types, such as cosmids and YACs, can be constructed and an optimal strategy of mapping of different resolutions, based on previously established PTSs, must be elaborated.

The high repeat frequency of the chromosome III and close resemblance between its telomeric sequences (effectively closing it into a circle), plus the higher number of clones hybridizing to probes mapped to other two chromosomes, demonstrate the possible problems with all types of algorithm with more complex genomes.

Case Study 16.1 (Continued)

40–50% chimaerism, which makes it difficult to build long-range contigs [35]. High chimaerism may affect single-chromosome projects to a lesser degree, but inconsistencies can still arise from deletions and internal rearrangements of large YAC clones. Estimates of chimaerism in large YACs vary from 40% [22] to 59% [36] to 80% [35]. To increase the reliability of resulting contigs, chimaeric clones often have to be excluded from the analysis, which is especially undesirable if the library redundancy is low.

In this section we briefly describe an approach that identifies both false positive signals and clones containing chimaeric inserts/internal deletions (see ref. 37 for fuller details). 'Dechimaerized' inserts are then represented as several independent contiguous clones, yielding a more consistent data set which may be ordered using existing tools. These inserts are referred to as 'components' of the original clone and can be later checked by other experimental methods to determine either the precise sites of coligation/deletion or the contents of a potential well contaminant.

The algorithm can be summarized as follows. Suppose a clone, C , is chimaeric, containing fragments from different regions of the genome. Then the set of single-copy probes that hybridize with C , P_C , divides into two or more groups, corresponding to those regions. If we can determine these subsets then C can be split into its components and we have effectively solved the problem. In essence, the algorithm checks if the probes in P_C are still connected when the clone C is ignored.

Consider a probe p in P_C . The subset $Q(p)$ of P_C is defined as p plus any other probe q in P_C , provided that at least l clones other than C connect q with any other probe in $Q(p)$. $Q(p)$ corresponds to a component of C . The same procedure is repeated with the probes remaining in P_C , leading to creation of several components of the original clone. If all the probes lie in a single component then the clone is deemed to be non-chimaeric.

For best results, the algorithm is applied iteratively. On iteration n , only clones hit between 2 and $N_0 + n$ times are analysed, $n = N_0 \dots N - N_0$, N being the maximum number of hits per clone in the library, while $N_0 > 1$ is the number of hits per clone on the first iteration. Obviously, if on the first iteration N_0 is set being equal to N , all the clones are analysed simultaneously. Starting at $N_0 < N$ allows us to avoid clones with the highest numbers of probe hits (as they are more likely to introduce non-linearities in the map) on early iterations and establish the most reliable links first. Both N_0 and l may be varied to obtain the best performance for data sets with different library redundancy, probe saturation and noise level.

The algorithm uses a depth-first search to produce for each clone one or more groups of probes mutually linked by at least l additional clones at each iteration, and replace each original library clone by one or more components consisting of identified groups of positives. Clone components positive with one probe only are omitted, while those yielding more than two positives are retained. Each library clone is analysed only once.

In the simulation example below, the correct order of probes was known *a priori*, but this information was not available to the program **chimaera**, an implementation of the above algorithm. The resulting orders of clone components were produced using the program **reorder** and the graphic output by the program **show**.

We simulated a data set represented by a mapping project of a 90-Mb chromosome covered by 300 clones of average length 1.5 Mb (library redundancy, 5), hybridized with 200 SC probes (probe saturation, 3.33). In order to estimate the algorithm's performance, a high chimaerism of 50% was simulated,

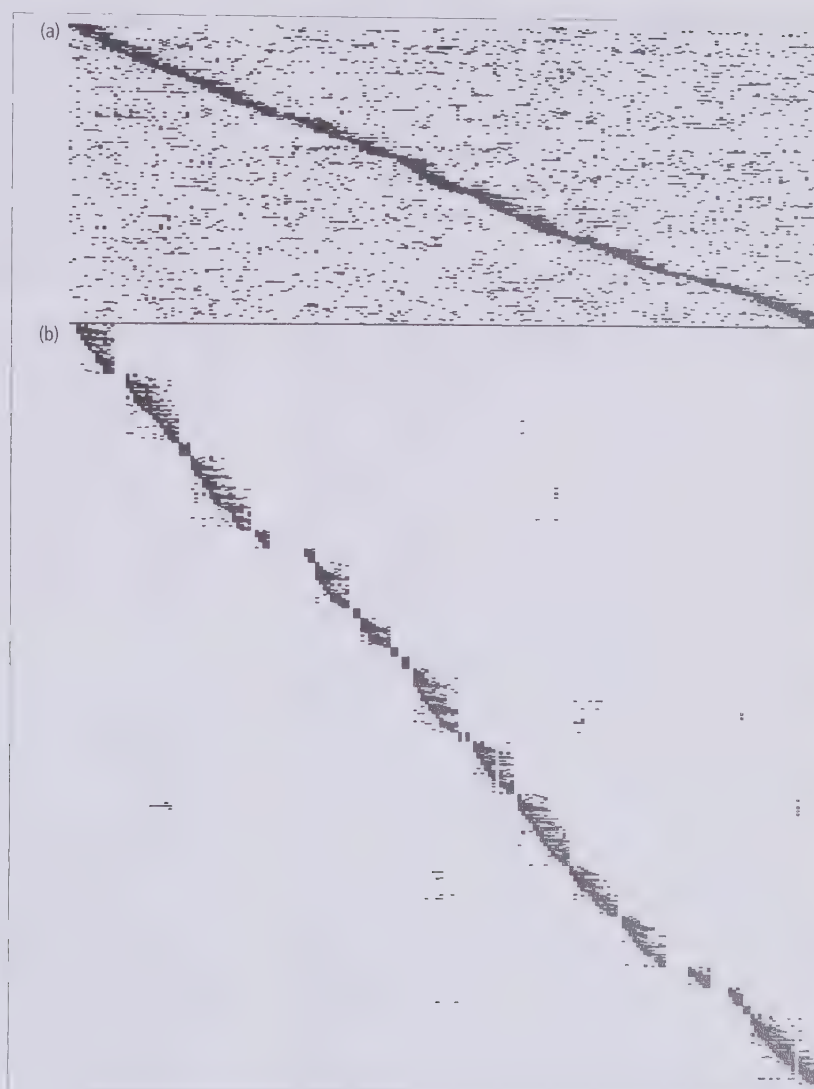


Fig. 16.3 Map of the simulated chimaeric data set. (a) Before and (b) after 'dechimaerization' by the program **chimaera** (see text).

accompanied by a false negative rate of 10%—that is, half of the library clones contained chimaeric inserts and one in 10 positive signals was not scored. In addition, a high rate of false positives (simulating nonspecific hybridization events or high over-scoring for a part of a library—for example, due to different filter-specific background levels) was introduced. All the clones contained false positives: for 60% of the library the false positive rate was set to 10%, and for the remaining 40% of clones the rate was set to 300%. That is, they yielded on average four times more random false positives. This data set is shown in Fig. 16.3a, where the 'veil' of noise almost hides the main diagonal.

The programs **probeorder** and **barr** [9] applied to this raw data set were unable to produce the correct map. **barr** eliminated most of the clones from the analysis and arrived at a large number of correct short contigs on average containing two probes, while **probeorder** produced two contigs containing 35

randomly ordered small groups of true neighbours.

The best results were obtained running the program **chimaera** with the parameters $N_0=16$ and $l=3$, when a 2.5-fold increase in the number of clones was detected. The resulting data set is given in Fig. 16.3b. Both **probeorder** and **barr** were able to construct 20 contigs, clearly visible as distinct groups of probes in Fig. 16.3b. Thirty-five probes (17.5%) appear in this figure as blank vertical lines because the resulting clone components positive with them were singletons (negative with the rest of the probes). These clone components provide no information for ordering probes and are omitted in the output.

16.8 Predictions of experimental progress in genomic mapping by anchoring random clones

We finish this chapter with the simplified mathe-

mathematical results of an analysis of the expected progress of a random anchor mapping project. The approach of linking random clones by hybridization with short anchor sequences has been the subject of intensive theoretical study [38–42]. Mostly following the style of ref. 43, which presented a mathematical analysis for physical mapping by fingerprinting random clones, these papers give equations representing properties of groups of anchored clones ('contigs', or anchored 'islands' of clones) obtained using anchor-based mapping.

The results of such an analysis can be used to help design and plan physical mapping projects using various types of anchors (e.g. restriction fragment length polymorphisms [44]; random amplified polymorphic DNAs, also known as arbitrarily primed PCR products [45,46]; short PCR assays for unique regions of the genome that have been dubbed sequence-tagged sites, applied to nested sets of clones [47]), all of which should obey the same rules.

Here, we summarize the published mathematical results and simplify them to a form suitable for non-mathematical readers. We also compare the corresponding expressions from all these papers to each other and compare the theoretical predictions with the results obtained in practice in constructing a complete physical map of YAC clones for the fission yeast *S. pombe* [1].

The efficiency of an anchor-orientated approach decreases markedly as more anchors are used, so it is necessary to change from the random mapping strategy and exploit other methods to bridge the remaining gaps between the contigs, as done with *S. pombe*. However, such a decision requires additional information about the number of gaps and undetected overlaps between islands at a given stage of the experiment, while the easiest accessible measures of the actual mapping progress are the number of anchored islands and singly anchored islands of clones (singletons). It has been shown [42] that the expected number of gaps can be approximated by the number of singleton anchored islands and the expected number of undetected overlaps between islands—by the number of islands containing more than one anchor. This proximity allows one to estimate mapping progress easily and decide on switching to (the more efficient) directed bridging of remaining gaps between contigs, while an estimation of other measures of progress (like length of anchored islands) may require additional experimental effort.

16.8.1 Notation and definitions

We define the following symbols:

G , haploid genome length in base pairs;

L , length of clone insert in base pairs;

N , number of clones in library;

M , number of anchors hybridized to clones;

$a = LN/G$, redundancy of coverage in clones or expected number of clones covering a random base pair;

$b = LM/G$, redundancy of coverage in anchors or expected number of anchors contained in a random clone.

An *anchored island* is a group of one or more clones linked together by anchors they share; an island formed by only one anchor is called a *singleton anchored island*. Clones on the ends of adjacent islands, however, can actually overlap but may not have any anchors in common, thus resulting in breaks in contigs. This case is referred to as an *undetected overlap* between the respective pair of islands. An *ocean* is a segment of the genome with no anchored islands on it.

16.8.2 Observations of the progress of mapping experiments

Here we compare the properties of contigs described by eqns 1–12 in Section 16.8.3 with the actual results obtained in the course of the physical mapping of *S. pombe* [1]. Anchors were taken randomly until b reached the value of 3. A set of 65 ordered YAC clones was then hybridized to the cosmid library and further cosmid anchors were selected among those not yet hit by the YAC clones. Also, the insert ends of the YAC clones positioned at the ends of contigs were used as anchor probes [1].

Four measures of the experimental progress are shown in Figs 16.4 and 16.5—the numbers of:

- 1 (all) anchored islands;
- 2 singleton anchored islands;
- 3 oceans;
- 4 undetected overlaps between anchored islands.

The left vertical axes give these values in units of G/L , making the graphs independent of sizes of the genome and clones, while the right vertical axes show the numbers observed for the *Schiz. pombe* mapping project.

The observed number of contigs obtained in the course of the 'random part' of this project are plotted in Fig. 16.4 together with the corresponding predicted values. There is a reasonable agreement between the theory and the actual experimental data, when one takes into account that the observed number of anchored islands must be integral and that the predictions ignore the fact that the genome of *Schiz. pombe* is divided into three chromosomes.

The difference between the predicted and ob-

served numbers of anchored islands and singleton anchored islands in the points of extremum in Fig. 16.4 does not exceed one island. The inflection to a slower decrease of the number of anchored islands for $b > 2$ for the theoretical curve together with the elimination of singleton anchored islands and oceans indicates the need to change the strategy of hybridization with anchors taken at random to directed bridging of remaining gaps representing undetected overlaps.

The numbers of oceans and undetected overlaps between islands in the course of the project are

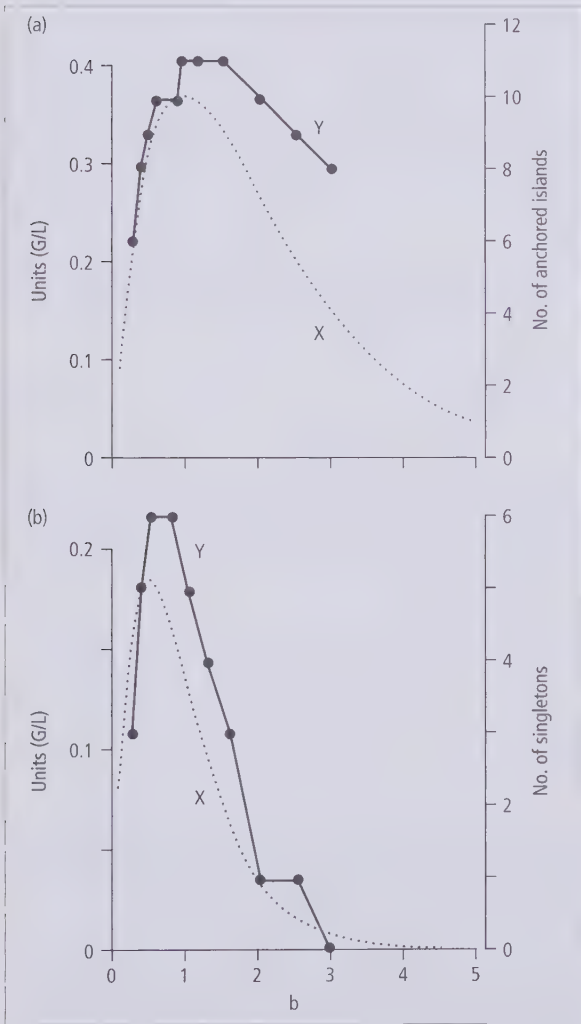


Fig. 16.4 Measures of experimental progress: the numbers of anchored islands and singleton anchored islands. (a) The expected (X) and experimentally observed (Y) numbers of anchored islands as a function of coverage b in anchors (see Equation 16.1, Section 16.8.3.2). (b) The expected (X) and experimentally observed (Y) numbers of singleton anchored islands as a function of coverage b in anchors (see Equation 16.2 in Section 16.8.3.2).

plotted in Fig. 16.5 together with the corresponding theoretical predictions, also in a good agreement. Notably, singleton anchored islands and oceans both disappear at $b = 3$. Two 'gaps' between chromosomes obviously represent the difference between the number of anchored islands in Fig. 16.4a and the sum of numbers of oceans and undetected overlaps in Fig. 16.5.

In Figs 16.4 and 16.5, both theoretical and experimental plots highlight four critical points for the anchor density b (and corresponding time spent on hybridizations):

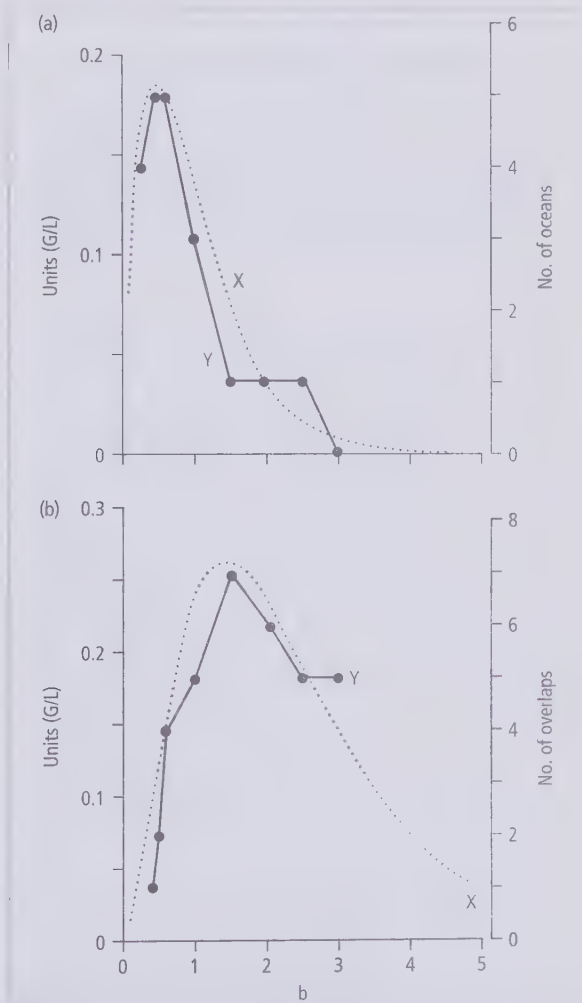


Fig. 16.5 Measures of experimental progress: the numbers of oceans and overlaps. (a) The expected (X) and experimentally observed (Y) numbers of oceans between anchored islands as a function of coverage b in anchors (see Equation 16.7 in Section 16.8.3.2). (b) The expected (X) and experimentally observed (Y) numbers of undetected overlaps between anchored islands as a function of coverage b in anchors (see Equation 16.9 in Section 16.8.3.2).

$b=0.5$: maximum number of singleton anchored islands and maximum number of oceans between islands;

$b=1$: maximum number of anchored islands;

$b=1.5$: maximum number of undetected overlaps between anchored islands;

$b=3$: elimination of singleton anchored islands and oceans.

Elimination of singleton anchored islands and oceans will take place later in mapping projects with a higher G/L ratio but the four plotted measures of the experimental progress must behave similarly to the theoretical curves shown in Figs 16.4 and 16.5.

16.8.3 Simplification of theoretical predictions

16.8.3.1 Practical considerations and assumptions

The equations obtained in refs 38–41 assume that N and M , and, hence, a and b , vary independently in the domains $[0, \infty]$. However, simple practical considerations allow for some narrowing of these domains.

First, it seems unreasonable to start a serious mapping project with a clone library equivalent to less than 4.61 genomes, as then a probability that a piece of genome is not cloned is more than $e^{-4.61} = 0.01$. Difficulties in cloning specific genomic regions, chimaerism of library clones and experimental noise stand in favour of at least doubling this redundancy in order to complete the project.

Second, the efficiency of an anchor-orientated mapping approach will decrease as more anchors are added, so it may be necessary to change from the random mapping technique and exploit other methods to bridge the final gaps between contigs obtained so far. This means that the number of random anchors may be small compared to the number of clones and nonrandom probing techniques should play an equally important part in completing the map.

Hence, in practice we may assume:

1 the number of anchors used being such that $b/a \ll 1$;

2 a redundancy of coverage a such that $e \gg 1$.

Under these assumptions, the equations from refs 38–41 can be approximated by much simpler expressions. Here, we also assume that all the clones in a library have the same length. This case was analysed in refs 38–41 and this allows us to compare their theoretical predictions to each other. However, given some distribution of clone lengths one can easily perform the simplifications analogous to those presented below using the assumptions 1 and 2 for any general-form equations in refs 38–41.

16.8.3.2 Number of islands

Reference 38 gives a definition of an island of clones different from that in ref. 39, in that an unanchored clone is to be regarded as an island as well. However, the expected number of islands of single-copy landmarks (SCLs) and the expected number of isolated SCLs estimated in ref. 38 are equivalent to the expected number of islands of anchored clones and expected number of singleton anchored islands, respectively, calculated in the other papers.

Despite minor differences, the equations for the expected number of anchored islands from all three papers simplify under assumptions 1 and 2 to the same expression:

$$Me^{-b} \quad (16.1)$$

And the corresponding equations of the expected number of singleton anchored islands from refs 38 and 40 result in

$$Me^{-2b} \quad (16.2)$$

Obviously, the expected number of non-singleton anchored islands will be

$$Me^{-b} - Me^{-2b} \quad (16.3)$$

16.8.3.3 Length of an island

Analogous simplifications of the formulae for the expected length of an anchored island result in

$$L \frac{e^b - 1 - b}{b} + 2L \left[1 - \frac{1}{a} \right] \quad (16.4)$$

where

$$2L \left[1 - \frac{1}{a} \right] \quad (16.5)$$

represent a span of the clones anchored on both ends of an island and is equivalent to the expected length of a singleton anchored island.

The discussion of anchor-based and clone-based sampling of islands in ref. 40 pointed out the relationship between bias in sampling of islands and the famous ‘waiting time’ paradox in probability. On average, an island containing a randomly chosen anchor will be larger than a randomly chosen island, as randomly chosen anchors will be more likely to fall in larger islands.

Starting from some anchor saturation, a new probe is most likely to be added to an existing contig, and thus only reduces the number of islands or additional coverage of the genome when hitting the

span of the clones anchored on the ends of that contig. Given the actual value of a , one can estimate from Equation 16.4 the anchor density b when the expected distance between rightmost and leftmost anchors of an island exceeds this span.

16.8.3.4 Proportion of genome not covered by anchored islands

Simplification of the corresponding formulae from all papers [38–41] produces the same result

$$e^{-2b} \tag{16.6}$$

16.8.3.5 Number and size of oceans and overlaps

In ref. 40, a formula is given for the probability that an island is followed by an ‘actual’ ocean—that is, a stretch of the genome not represented in the library and thus having no clone (either anchored or not anchored) on it. This was extended in ref. 39 to all possible types of oceans permitting the calculation of the expected number and size of oceans and overlaps between islands.

Simplifying corresponding equations from ref. 39 for the expected number of oceans we get:

$$N \frac{e^b - 1 - b}{e^{a+b}} + Me^{-2b} \tag{16.7'}$$

and for the expected ocean size

$$\frac{L}{b+a} \frac{e^b - 1 - b}{e^{a+b}} \tag{16.8'}$$

Also from ref. 39, for the expected number of overlaps between islands we obtain

$$Me^{-b} - Me^{-2b} - N \frac{e^b - 1 - b}{e^{a+b}} \tag{16.9'}$$

and for the expected overlap length

$$L \frac{b + e^{-b} - 1}{b(1 - e^{-b}) - a} \frac{e^b - 1 - b}{e^a} \tag{16.10'}$$

These expressions (7'–12') may be simplified further for large a . In particular, for the expected number of oceans we have

$$Me^{-2b} \tag{16.7}$$

and for the expected number of overlaps

$$Me^{-b} - Me^{-2b} \tag{16.9}$$

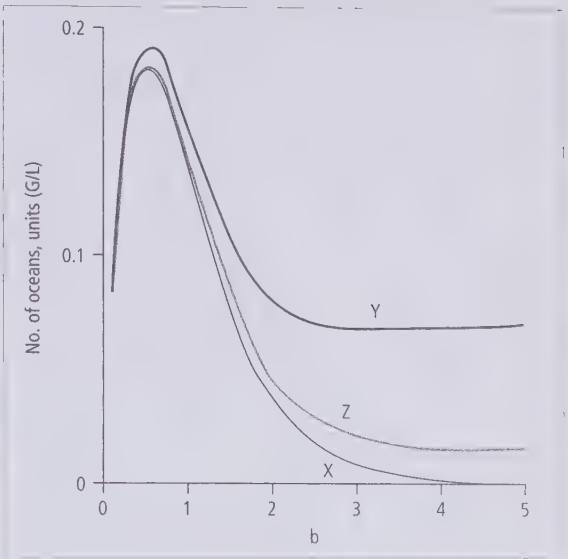


Fig. 16.6 The expected number of oceans between anchored islands as a function of coverage b in anchors according to Equation 16.7 (X) and Equation 16.7', for $a = 4$ (Y) and $a = 6$ (Z).

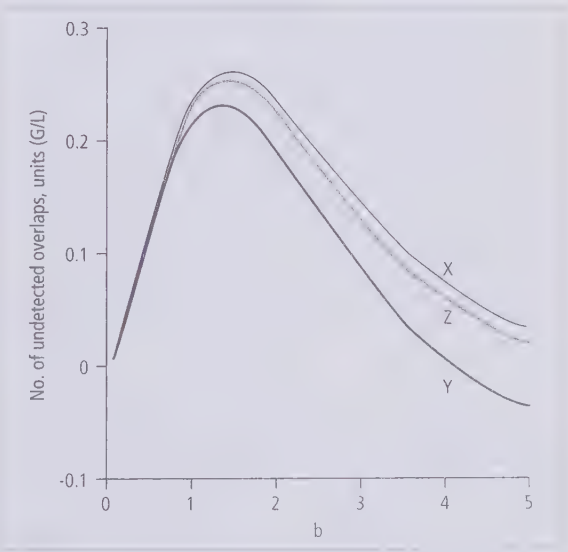


Fig. 16.7 The expected number of undetected overlaps between anchored islands as a function of coverage b in anchors according to Equation 16.9 (X) and Equation 16.9' for $a = 4$ (Y) and $a = 6$ (Z).

These expressions are identical to those obtained for the expected number of singleton anchored islands and the expected number of anchored islands with more than one anchor, respectively. This can also be seen from ref. 39, where, together with the calculations of ocean and overlap properties, numerical

results for the *Arabidopsis* genome mapping project are presented. Predicted values are:
the number of singleton anchored islands: 65.54
the number of oceans: 68.14
the number of nonsingleton anchored islands: 115.47
the number of overlaps: 112.87
for $a = 5.75$ and $b = 1.25$.

A possible explanation for such a proximity between the numbers of oceans and singleton anchored islands, and between numbers of undetected overlaps and nonsingleton anchored islands, respectively, is as follows. In early stages of mapping ($b < 0.5$), most anchored islands are singletons divided by oceans. Each new anchor is most likely to produce a new singleton, thus also increasing the number of oceans. Transformation from a singleton to a nonsingleton anchored island enlarges the span of anchored clones around it and eventually eliminates the adjacent oceans. Islands remaining singletons till the high anchor saturations are likely to fall into regions of a poor probe contents and lower clone densities and to consist of shorter clones, all factors in favour to 'preserve' oceans surrounding them. Plots in Figs 16.4b and 16.5a show that the number of singleton anchored islands can be used to estimate the number of oceans between islands.

In later stages ($b > 1.5$), nonsingleton islands dominate among contigs. Connection of two islands into one can take place only if there is an undetected overlap between them and a new anchor hits this overlap. Therefore, the number of undetected overlaps can be estimated by the number of nonsingleton anchored islands.

As to length characteristics of oceans and overlaps, for the expected size of an ocean we obtain

$$L/b \tag{16.8}$$

and for the expected length of an undetected overlap between islands

$$L \left[\frac{1}{1 - e^{-b}} - \frac{1}{b} \right] \tag{16.10}$$

for large a .

Contribution of the a -dependent components of expressions 16.7' and 16.9' is illustrated graphically in Figs 16.6 and 16.7, respectively. One can easily see that it is negligibly small for $a > 6$. Such a contribution is even smaller for expressions 16.8' and 16.10' (data not shown).

References

1 Maier, E., Hoheisel, J.H., McCarthy, L. *et al.* (1992)

Complete coverage of the *Schizosaccharomyces pombe* genome in yeast artificial chromosomes. *Nature Genet.* **1**, 273–277.
2 Hoheisel, J.D., Maier, E., Mott, R. *et al.* (1993) High-resolution cosmid and P1 maps spanning the 14-Mbp genome of the fission yeast *Schizosaccharomyces pombe*. *Cell* **73**, 1–20.
3 The Huntington's Disease Collaborative Research Group. (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983.
4 Baxendale, S., MacDonald, M.E., Mott, R. *et al.* (1993) A cosmid contig and high resolution restriction map of the 2 megabase region containing the Huntington's disease gene. *Nature Genet.* **4**, 181–186.
5 Hudson, T.J., Stein, C.D., Gerety, S.S. *et al.* (1995) An STS-based map of the human genome. *Science* **270**, 1945–1954.
6 Barillot, E., Lacroix, B. & Cohen, D. (1991) Theoretical analysis of library screening using an N-dimensional pooling strategy. *Nucleic Acids Res.* **19**, 6241–6247.
7 Mott, R., Grigoriev, A., Maier, E. *et al.* (1993) Algorithms and software tools for ordering clone libraries, application to the mapping of the genome of *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **21**, 1965–1974.
8 Kohara, Y., Akiyama, A. & Isono, K. (1987) The physical map of the whole *E. coli* chromosome, application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**, 495–508.
9 Olson, M.V., Dutchik, J.E., Graham, M.Y. *et al.* (1986) Random-clone strategy for genomic restriction mapping in yeast. *Proc. Natl. Acad. Sci. USA* **83**, 7826–7830.
10 Coulson, A., Sulston, J., Brenner, S. & Carn, J. (1986) Towards a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **83**, 7821–7825.
11 Sulston, J., Mallet, F., Staden, R. *et al.* (1988) Software for genome mapping by fingerprinting techniques. *Comp. Appl. Biosci.* **4**, 125–132.
12 Balding, D.J. & Torney, D.C. (1991) Statistical analysis of DNA fingerprint data for ordered clone physical mapping of human chromosomes. *Bull. Math. Biol.* **53**, 853–879.
13 Chumakov, I., Rigault, P., Guillou, S. *et al.* (1992) Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* **359**, 380–387.
14 Cox, D.R., Burmeister, M., Price, E.R. *et al.* (1990) Radiation hybrid mapping, a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250**, 245–250.
15 Nizetic, D., Gellen, L., Hamvas, R. *et al.* (1994) An integrated YAC-overlap and 'cosmid-pocket' map of the human chromosome 21. *Hum. Mol. Genet.* **3**, 759–770.
16 Lehrach, H., Drmanac, R., Hoheisel, J. *et al.* (1990) Hybridization fingerprinting in genome mapping and sequencing In *Genome Analysis. Vol. 1: Genetic and Physical Mapping* (Davies, K.E. & Tilghman, S.M., eds), pp. 35–81 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
17 Meier-Ewert, S., Maier, E., Ahmadi, A. *et al.* (1993) An automated approach to generating expressed sequence catalogues. *Nature* **361**, 375–376.

- 18 Boehnke, M., Lange, K. & Cox, D.R. (1991) Statistical methods for multipoint radiation hybrid mapping. *Am. J. Hum. Genet.* **49**, 1174–1188.
- 19 Perlin, M. & Chakravarti, A. (1993) Efficient construction of high-resolution physical maps from yeast artificial chromosomes using radiation hybrids—inner-product mapping. *Genomics* **18**, 283–289.
- 20 Weissenbach, J., Gyapay, G., Dib, C. *et al.* (1992) A 2nd-generation linkage map of the human genome. *Nature* **359**, 749–801.
- 21 Dietrich, W.F., Miller, J.C., Steen, R.G. *et al.* (1994) A genetic-map of the mouse with 4,006 simple sequence length polymorphisms. *Nature Genet.* **7**, 220–245.
- 22 Bellanné-Chantelot, C., Lacroix, B., Ougen, P. *et al.* (1992) Mapping the whole human genome by fingerprinting yeast artificial chromosomes. *Cell* **70**, 1059–1068.
- 23 Cohen, D., Chumakov, I. & Weissenbach, J. (1993) A first-generation physical map of the human genome. *Nature* **366**, 698–701.
- 24 Breen, M., Deakin, L., MacDonald, B. *et al.* (1994) Towards high-resolution maps of the mouse and human genomes—a facility for ordering markers to 0.1 cM resolution. *Hum. Mol. Genet.* **3**, 621–627.
- 25 Cox, R.D., Copeland, N.G., Jenkins, N.A. & Lehrach, H. (1991) Interspersed repetitive element polymerase chain reaction product mapping using a mouse interspecific backcross. *Genomics* **10**, 375–384.
- 26 McCarthy, L., Hunter, K., Schalkwyk, L. *et al.* (1995) Efficient high-resolution genetic mapping of mouse interspersed repetitive sequence PCR products, toward integrated genetic and physical mapping of the mouse genome. *Proc. Natl. Acad. Sci. USA* **92**, 5302–5306.
- 27 Durbin, R. & Thierry Mieg, J. (1991) A C. elegans Database. Documentation, code and data available from anonymous ftp servers at lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk and ncbi.nlm.nih.gov.
- 28 Cuticchia, A.J., Arnold, J. & Timberlake, W.E. (1993) Ods—ordering DNA-sequences—a physical mapping algorithm based on simulated annealing. *Comp. Appl. Biosci.* **9**, 215–219.
- 29 Wang, Y.H., Prade, R.A., Griffith, J. *et al.* (1994) A fast random cost algorithm for physical mapping. *Proc. Natl. Acad. Sci. USA* **91**, 11094–11098.
- 30 Newberg, L.A. (1994) Finding a most likely clone ordering from oligonucleotide hybridization data. *Genomics* **21**, 602–611.
- 31 Zhang, P.S., Schon, E.A., Fischer, S.G. *et al.* (1994) An algorithm based on graph theory for the assembly of contigs in physical mapping of DNA. *Comp. Appl. Biosci.* **10**, 309–317.
- 32 Clark, D.A., Rawlings, C.J. & Doursenot, S. (1994) Genetic map construction with constraints. In *Proceedings of the Second International Conference on Intelligent Systems in Molecular Biology* (AAAI/MIT Press, Menlo Park, CA), pp. 78–86.
- 33 Metropolis, N., Rosenbluth, A., Rosenbluth, M. *et al.* (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092.
- 34 Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. (1989) *Numerical Recipes in C* (Cambridge University Press, Cambridge).
- 35 Andersson, C. (1993) Genome shortcut leads to problems. *Science* **259**, 1684–1687.
- 36 Foote, S., Vollrath, D., Hilton, A. & Page, D.C. (1992) The human Y chromosome, Overlapping DNA clones spanning the euchromatic region. *Science* **258**, 60–66.
- 37 Grigoriev, A., Mott, R. & Lehrach, H. (1994) An algorithm to detect chimaeric clones and random noise in genomic mapping. *Genomics* **22**, 482–486.
- 38 Barillot, E., Dausset, J. & Cohen, D. (1991) Theoretical analysis of a physical mapping strategy using random single-copy landmarks. *Proc. Natl. Acad. Sci. USA* **88**, 3917–3921.
- 39 Ewens, W.J., Bell, C.J., Donnelly, P.J. *et al.* (1991) Genome mapping with anchored clones, theoretical aspects. *Genomics* **11**, 799–805.
- 40 Arratia, R., Lander, E.S., Tavare, S. & Waterman, M.S. (1991) Genomic mapping by anchoring random clones, a mathematical analysis. *Genomics* **11**, 806–827.
- 41 Torney, D.C. (1991) Mapping using unique sequences. *J. Mol. Biol.* **217**, 259–264.
- 42 Grigoriev, A. (1993) Theoretical predictions and experimental observations of genomic mapping by anchoring random clones. *Genomics* **15**, 311–316.
- 43 Lander, E.S. & Waterman, M.S. (1988) Genomic mapping by fingerprinting random clones, a mathematical analysis. *Genomics* **2**, 231–239.
- 44 Chang, C., Bowman, J.L., DeJohn, A.W. *et al.* (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **85**, 6856–6860.
- 45 Williams, J.G.K., Kubelik, A.R., Livak, K.J. *et al.* (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**, 6531–6535.
- 46 Welsh, J. & McClland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* **18**, 7213–7218.
- 47 Green, E.D. & Olson, M.V. (1990) Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes, a model for human genome mapping. *Science* **250**, 94–98.
- 48 Fan, J.-B., Chikashige, Y., Smith, C.L. *et al.* (1988) Construction of a Not I restriction map of the fission yeast *Schizosaccharomyces pombe* genome. *Nucleic Acids Res.* **17**, 2801–2818.

Chapter 17 Strategies for rapid isolation of genes from genomic DNA: cDNA techniques

Michael A. North, Bernhard Korn*, Roger Cox** & Hans Lehrach***

Sequana Therapeutics, 11099 North Torrey Pines Road, Suite 160, La Jolla CA 92037, USA

**ICRF, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK*

***ICRF, Wellcome Trust Centre for Human Genetics, Nuffield Department of Clinical Medicine, Windmill Road, Headington, Oxford OX3 7BN, UK*

****Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin, Germany*

17.1	Introduction, 442
17.1.1	Exon trapping, 442
17.1.2	cDNA enrichment, 443
17.1.3	Exon amplification and cDNA enrichment as complementary approaches, 443
17.1.4	Alternative strategies for isolating coding sequence, 444
17.2	Exon trapping by pSPL1, 445
17.2.1	Efficiency and specificity of exon amplification, 445
17.2.2	Artefacts, 446
17.2.3	Chimaeric exons, 446
17.2.4	Modifications to exon-trapping protocol, 447
17.2.4.1	Isolating expressed sequences from YACs, 447
17.2.4.2	The pSPL3 exon-trapping vector, 447
17.3	Direct cDNA hybridization methods, 447
17.4	cDNA enrichment, 447
17.5	Obtaining a full-length cDNA, 448
17.5.1	3' RACE, 448
17.5.2	5' RACE, 448

Protocol 91
Exon trapping by pSPL1 from genomic DNA cloned in cosmids, 449
Troubleshooting, 455
Ratio of recombinants to non-recombinants is less than 10:1, 455
Protocol 92
Direct cosmid or YAC hybridization to cDNA filters, 456
Protocol 93
cDNA enrichment using solution hybridization to genomic clones as probes, 458
Protocol 94
Rapid amplification of 3' ends, 462
Protocol 95
5' RACE using a unique oligonucleotide ligated to the 3' end of the first-strand cDNA, 464
References, 466

17.1 Introduction

The aim of positional cloning is to identify and isolate the coding sequences within the genomic region of interest, one of which will be the sought-for gene. This ‘endgame’ can be the rate-limiting step in situations where the minimal genetic interval corresponds to several megabases of genomic DNA. There are many different strategies for isolating coding sequence from genomic DNA and this chapter will cover two of the most important and widely used—*exon trapping* and *cDNA enrichment*—which are complementary in approach and technically straightforward. Also covered here are methods for obtaining full-length cDNAs. Further characterization of transcripts in terms of expression profiles, mutational analysis and functional studies can be carried out using methodologies described elsewhere [1].

17.1.1 Exon trapping

Exon-trapping strategies have now been used successfully in many positional cloning projects (e.g. for the Menkes disease gene [2], the neurofibromatosis type 2 tumour suppressor gene [3], the

Huntington’s disease gene [4], etc.). Many different exon-trapping protocols have been published [5–8]. Protocol 91 describes in detail exon isolation from single cosmids with the exon-trapping vector pSPL1 [7], available from Gibco-BRL. This system is presently the most widely used and extensively characterized of the exon-trapping protocols. pSPL1 is a mammalian expression vector that has been engineered to allow genomic DNA to be inserted into an intron flanked by the 5’ and 3’ splice sites of the human immunodeficiency virus (HIV-1) *tat* gene. Recombinant clones are transfected into COS-7 cells and high levels of transcription are driven by the expression vector’s SV40 early promoter. During *in vivo* processing of transcripts, the splice sites of any exon contained within an inserted genomic fragment are paired with the *tat* splice sites so that intronic DNA is excised and the exon is retained in the mature RNA. Reverse transcription followed by polymerase chain reaction (PCR) can then be used to amplify such ‘trapped’ exons. A schematic of this method is shown in Fig. 17.1. If a cloned DNA fragment does not contain an exon, all the cloned DNA is spliced out of the primary transcript along with the surrounding vector intron sequences, yielding an mRNA containing only pSPL1-derived

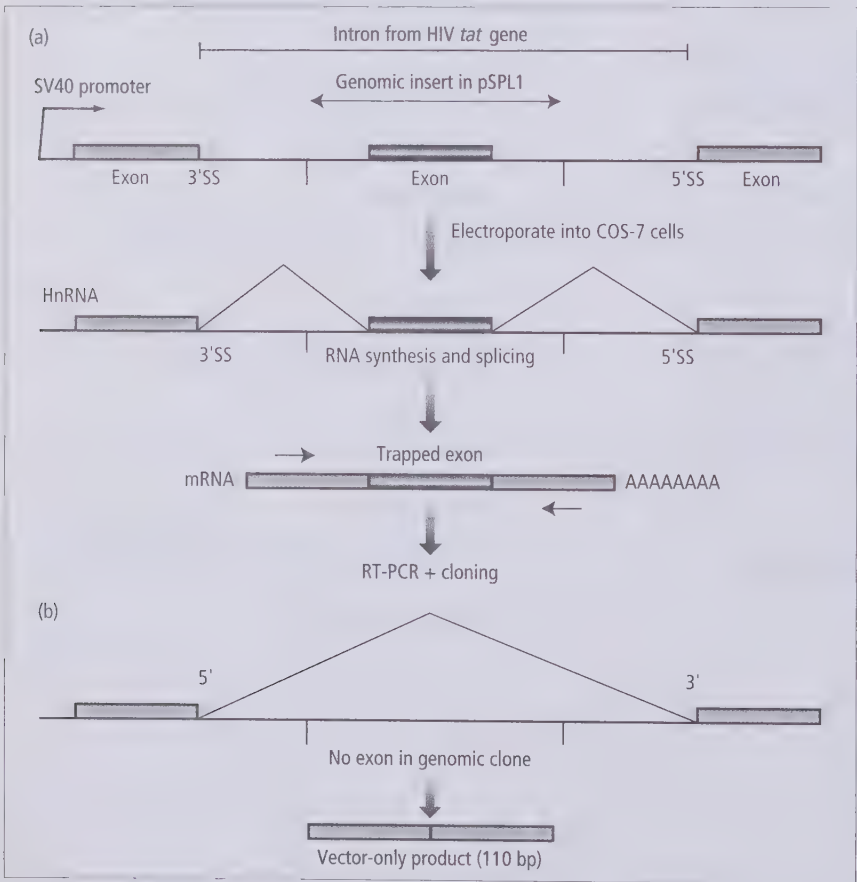


Fig. 17.1 Schematic illustration of the pSPL1 exon-trapping system. (a) A fragment of genomic DNA containing an exon is subcloned into the pSPL1 plasmid and electroporated into COS-7 cells. After transcription, vector-derived 5’ and 3’ splice junctions (5’Ss and 3’Ss) pair with the sequences flanking a cloned exon, removing intervening noncoding sequence by splicing. The trapped exon can then be amplified by reverse transcription followed by PCR (RT-PCR). (b) If the genomic fragment lacks an exon, it is spliced out of the hnRNA completely.

sequence. The basic protocol can be modified in several ways, which are discussed later. Exons isolated by this method can be subcloned and sequenced and used as hybridization probes on cDNA libraries to obtain full-length cDNA clones [1].

17.1.2 cDNA enrichment

The simplest method of screening for coding sequences is to hybridize genomic DNA directly to cDNA library filters. Genomic hybridization probes are most successful if the genomic fragment is relatively small and repeat-free, and this can represent a very direct strategy for gene isolation [9–11]. Hybridization of cosmids or yeast artificial chromosomes (YACs) containing genomic inserts directly to cDNA library filters is possible but is technically more difficult, mainly due to cross-hybridization of low and medium copy number repeats and the small proportion of the probe corresponding to coding sequence (low signal-to-noise ratio). A large YAC may detect a large number of cDNAs within a library, and many of these may be false positives containing repeats (in one controlled study a false positive background of around 90% was observed [12]). Hence an adequate screening system for clones of interest should be considered.

A modification of this very basic procedure, often referred to as *cDNA hybrid selection* or *cDNA enrichment*, is generally more successful in identifying cDNA using whole cosmids or YACs as probes. Several strategies have been described for optimizing the hybridization of YAC DNA to cDNA either in solution or with the YAC bound to a solid support [13–17]. The method described in detail in this chapter (Protocol 92) is one of the most technically straightforward, in our hands successfully isolates locus-specific transcripts, and has been tested rigorously in a controlled study [17]. In this method solution hybridization is typically carried out between a biotinylated YAC or cosmid and a PCR-amplified cDNA library. cDNAs specifically hybridizing to a particular genomic DNA sequence are selected by a biotin–streptavidin interaction (using streptavidin-coated magnetic beads) and the nonspecific hybrids are dissociated by stringent washing. The cDNAs selected are eluted, amplified and cloned, and comprise a ‘region-specific sub-library’ of the total cDNA isolated from a particular tissue or mixture of tissues. The technique results in an enrichment of the selected cDNAs of between 10^3 and 10^6 . It allows the simultaneous analysis of several large genomic intervals of varying complexities, and can be used to isolate the same expressed sequences from different tissues in parallel.

Any cDNA hybridization strategy should take into consideration that only some 10–20% of all mRNAs may be expressed in any differentiated cell type (about 10 000 genes per cell type). The level of expression of these genes may be as high as 200 000 mRNA molecules per single cell or as low as <1 molecule per cell, with $\approx 30\%$ of the genes expressed at <10 copies per cell at any given time in cellular development [18,19]. Therefore, a few million clones at least must be screened to have a reasonable chance of finding a particular low-abundance transcript. The search is further complicated when using a complex tissue as a source for cDNA as different cell types are present, each containing varying transcript abundance classes, decreasing the representation of cell-type specific abundance classes [20,21]. It is unlikely that a conventional cDNA library of a few million clones will adequately represent all transcripts that are expressed in a given tissue, because it will probably not contain low-abundance transcripts. Recently, a number of approaches to overcoming these problems have been suggested, and strategies have been designed to normalize the cDNA abundance classes by reassociating and removing the abundant cDNAs so that fewer clones need be screened [22].

17.1.3 Exon amplification and cDNA enrichment as complementary approaches

Both exon amplification and cDNA enrichment can be used on their own but a particularly efficient way of screening large genomic regions for genes is to use these two techniques in parallel. For example, a set of cosmids spanning a genomic region can first be used in the cDNA enrichment protocol (Protocol 93) to generate a *minilibrary* of cDNAs which are picked and stored in microtitre plates and spotted onto hybridization membranes. Clones containing repetitive elements can be identified by hybridization of Cot-1 DNA (see Protocols 92 and 93; see also Chapter 15, Protocol 89). The genomic cosmids are then exon trapped either singly or in pools and individual exon amplification products cloned. The set of candidate exons are then hybridized to the minilibrary membrane and the cDNAs hybridizing with each exon amplification product are recorded.

Typically, this procedure will identify a minimal set of cDNAs mapping to the region since—for example, different exons from the same gene detect overlapping sets of cDNAs in the minilibrary. *cDNA walking* to isolate a full-length transcript can be performed extremely rapidly within the enriched minilibrary. Examples of this methodology are shown in Fig. 17.2. Hybridization of even very small

exons to such library filters is very straightforward and the randomly primed cDNAs detected typically fall in the 500–1000 bp size range. The minilibrary array can be additionally hybridized with protein motif oligonucleotides, trinucleotide repeats, single-copy genomic fragments, etc. All the information from such additional screens can be easily integrated and a minimal set of cDNAs chosen for sequencing. It is often advantageous to sequence a minimal set of exon amplification products as well.

An important advantage of this approach is that it eliminates two common causes of difficulties with these techniques: (i) the identification of a full-length cDNA starting from a single exon amplification product and (ii) the complexity of the cloned product from cDNA enrichment. Also, artefacts such as pseudogenes (potentially isolated by cDNA enrichment) are generally not isolated by exon amplification.

17.1.4 Alternative strategies for isolating coding sequence

There are many other ways of isolating coding sequence from genomic DNA. For example, genomic clones can be tested for the presence of CpG islands by digestion with restriction enzymes whose recognition sequences include the CpG dinucleotide and which only cut if the CpG dinucleotide is not methylated (e.g. *NotI*, *SalI*, *SfiI*). Undermethylated CpG dinucleotides are associated with the 5' ends of some genes, particularly genes which are ubiquitously expressed [23]. Markers for the 5' end of some transcripts present in a region can be identified in this way, although actual identification of coding sequence will generally require additional work—for example, hybridizing single-copy probes from the region to zoo blots or cDNA libraries [24].

Another approach takes advantage of the fact that

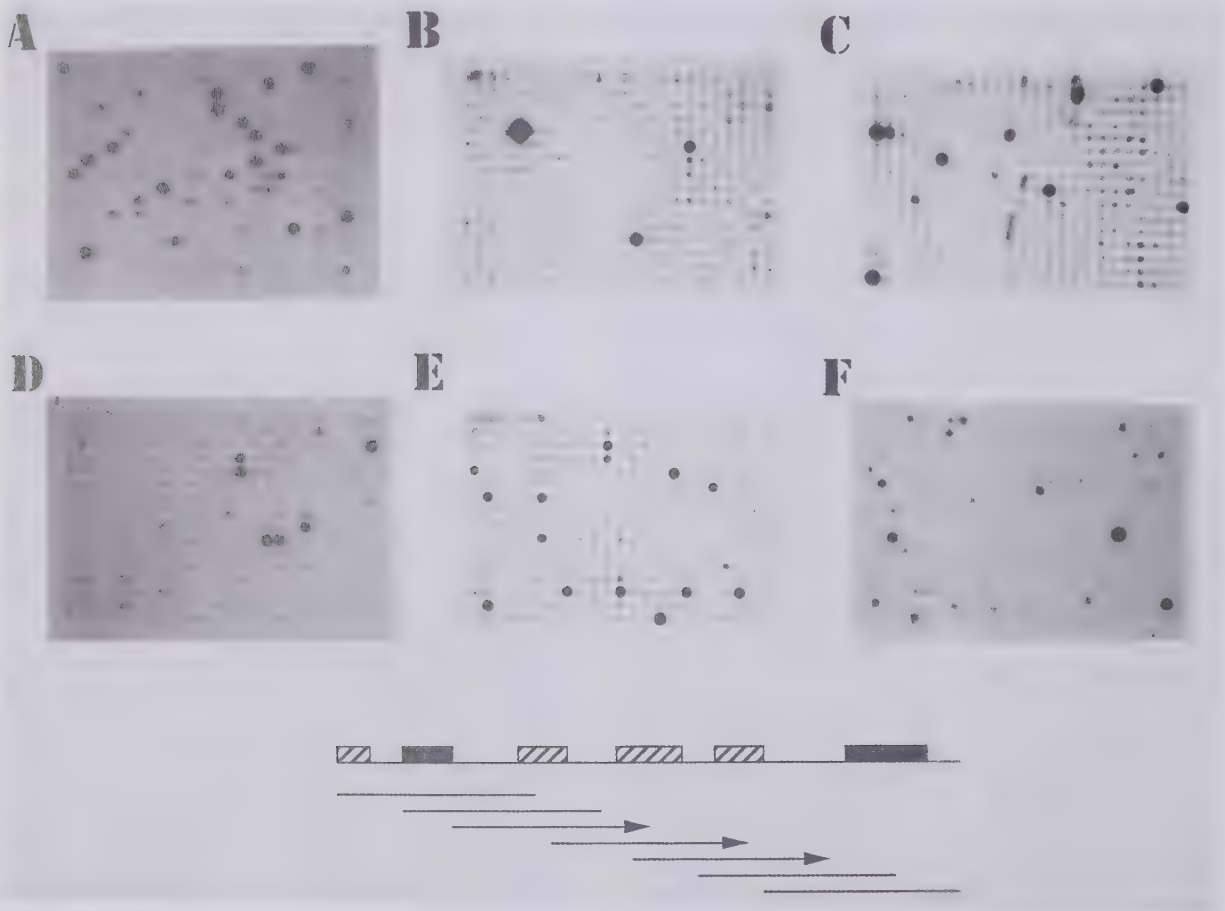


Fig.17.2 Examples of hybridizations of exon amplification products to gridded arrays of chromosome region-specific human cDNA minilibraries (prepared by enrichment methods). (A, F) Cot-1 hybridizations; (B) hybridization of *PPY*, a gene known to be present in the

region; (A, C, D, E) exon hybridizations. The schematic illustrates cDNA walking within the minilibrary between cDNAs detected by the hybridization of different exons from the same gene (solid blocks, exons isolated by trapping).

about 10% of human cDNAs contain species-specific repeats in the untranslated regions. Somatic cell hybrids (see Chapter 14) are generated containing a human chromosomal region of interest and the RNA isolated [25,26]. Then, for example, a cDNA library is constructed and screened with human repetitive DNA, and positively hybridizing cDNAs are then characterized further. A general limitation of these types of strategy is that the gene is expressed in the hybrid and that it contains human-specific repetitive sequence. Another alternative strategy detects coding sequence by homologous recombination between genomic clones and cDNA clones in a suitable genetic screen [27]. The genomic fragments used must be fairly small and free of repetitive elements; hence the strategy has no significant advantages over the direct cDNA hybridization approaches described above.

An important long-term consideration is that the genome mapping projects now in progress are, in the future, likely to provide large numbers of candidate genes, as transcriptional maps are integrated with physical maps. For example, the gene responsible for glycerol kinase deficiency (an X-linked inherited disease) has recently been identified both by a direct approach, in which transcripts in the appropriate genetic interval were examined, and by an indirect approach, in which large numbers of cDNAs were sequenced and one cDNA mapping to the X chromosome and with homology to a bacterial glycerol kinase was identified [28,29]. As increasing numbers of cDNAs or exons are

sequenced and mapped to chromosomes, this type of indirect approach will become increasingly important in positional cloning strategies as will the ability to access and query expressed sequence databases [30]. Additionally, computational methods for identifying coding regions in large stretches of sequenced genomic DNA will be increasingly useful to positional cloners as the genome sequencing project progresses [31].

17.2 Exon trapping by pSPL1

Protocol 91 describes exon trapping by pSPL1 from genomic DNA cloned in cosmids.

17.2.1 Efficiency and specificity of exon amplification

This method of rapid gene isolation is now well characterized. In one study, a 185-kb region of the human MHC class II region containing eight known genes was tested using the exon amplification system described in Protocol 91 [34]. Exons were recovered from seven out of eight known genes and two new expressed sequences were identified. The one known gene that was not detected is entirely contained on a large (20 kb) *Bam*HI/*Bgl*III fragment and was very inefficiently cloned into the pSPL1 vector. Repeating the experiment using a partial *Sau*3 A digest of the appropriate cosmid avoided this problem. As illustrated in Fig. 17.3b, the pattern of trapped exons can vary substantially, depending on

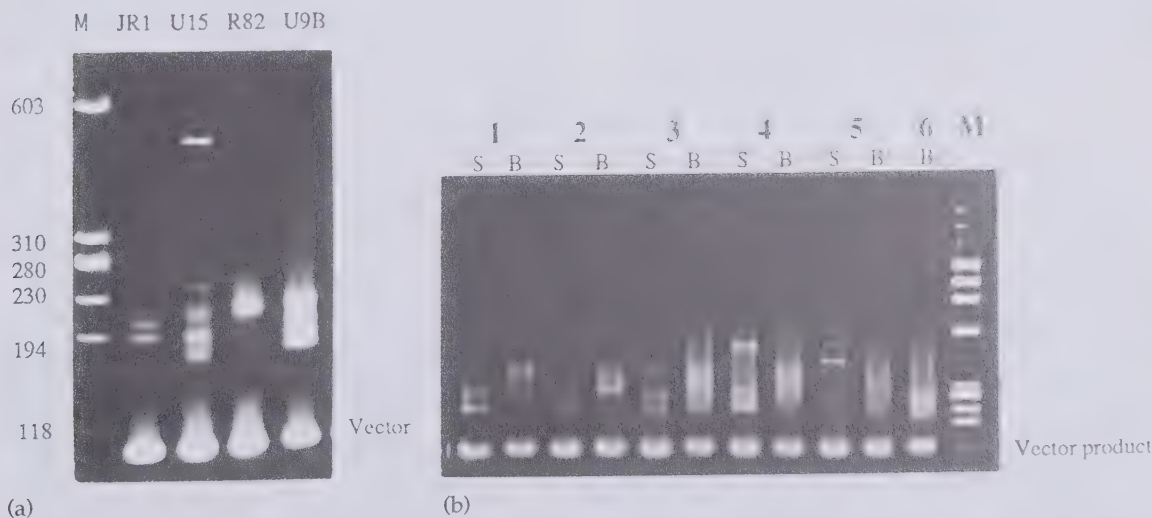


Fig. 17.3 Examples of pSPL1 exon amplification from cosmids. (a) pSPL1 exon amplification from cosmids JR1, U15, R82 and U9B. Products of size greater than the vector band correspond to trapped exons. Size markers (M) are indicated in bp. (b) Comparison of exon

amplification product patterns obtained by digestion of pools of 10 cosmids (1–6) with either *Bam*HI/*Bgl*III (B) or *Sau*3 A (S). PCR products of size greater than the indicated vector-only product correspond to trapped exons.

the choice of restriction enzyme and it is therefore advisable to use both *Sau3A* and *BamHI/BglII* cosmid digests in an exhaustive screen for exons. However, even in this simplest case of a *BamHI/BglII* genomic digest, the system can be estimated to detect the vast majority of genes in a genomic sample. It is important to remember, however, that genes that are comprised of only one or two exons (i.e. whose exons are not flanked by both donor and acceptor splice junctions) are, in principle, not detected by the system. The efficiency of gene isolation is not significantly compromised by using pools of 5–10 cosmids rather than trapping each cosmid individually [35].

17.2.2 Artefacts

Exon amplification will also produce a small number of PCR products arising from the amplification of noncoding sequences which do, however, contain regions with high homology to acceptor and donor splice junctions [34]. Examples of artefacts of this nature are illustrated in Fig. 17.4. For example, the Alu repeat shown contains regions that may act as 5' and 3' splice junctions so that the intervening 108 bp can be amplified. This particular Alu structure is presumably fairly rare in the genome since the vast majority of Alu repeats are not amplified by this system. Subfamilies of other medium and high copy number repeats such as O-ring repeats and the mouse LINE and HSAG repeats have occasionally also been amplified (M.N., unpublished). Figure 17.4 also illustrates a 241-bp amplification product derived purely from the pSPL1 *tat* intron following splicing directed by cryptic donor and acceptor junctions. Several other

products derived purely from the *tat* intron have also been observed. It is important to emphasize that artefactual amplifications such as those detailed above are fairly rare (about 15% of all amplified products). Artefacts seem to fall into a small number of categories and can be eliminated at an early stage of analysis. For example, the PCR products can be conveniently tested by hybridization to a *BamHI/BglII* digest of the original cosmid and to a Southern blot of appropriate genomic DNA. Amplification of repeat elements is immediately apparent following hybridization to genomic DNA while products derived purely from the *tat* intron do not hybridize to the cosmid digest.

17.2.3 Chimaeric exons

A relatively common event in this system is the isolation of PCR products containing both unique sequence and material derived from the *tat* intron [34]. This can occur when an exon is interrupted by a *BamHI* or *BglII* restriction site so that the exon is only amplified following compensation for the loss of the normal 5' splice junction by the activation of a cryptic 5' splice junction in the *tat* intron 66 bp downstream of the *BamHI* cloning site. The specificity of the system for amplification of exons does not appear to be significantly compromised by the loss of a splice junction. In many cases examined, the unique sequence either shows homology to a known gene or is conserved on zoo blots [34]. Activation of this cryptic splice junction may often occur only in the absence of a normal mammalian 5' splice junction and allows the 'rescue' of exons which would otherwise not be cloned in the experiment.

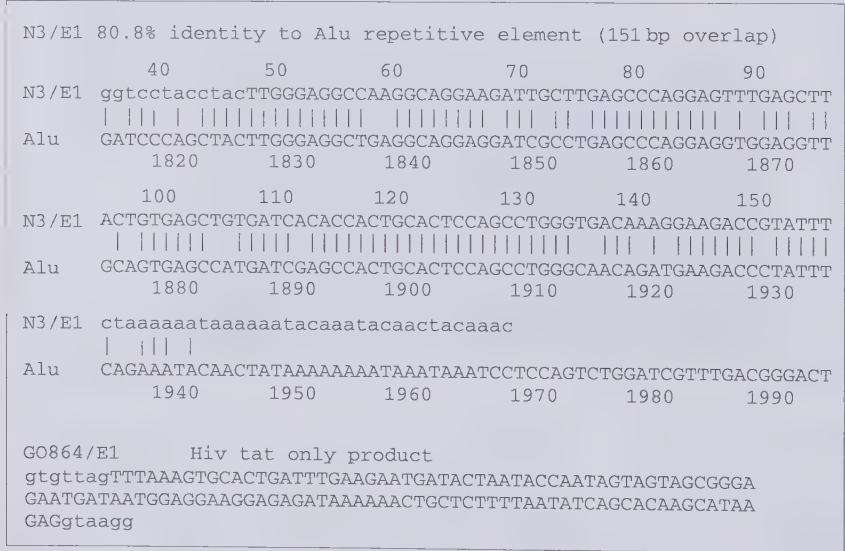


Fig. 17.4 Examples of an artefact amplification product. N3/E1 is an Alu repeat cloned from MHC cosmids U15 and N3. Lowercase nucleotides indicate flanking sequences which can act as splice junctions. GO864/E1 is a product derived solely from the HIV *tat* intron (again, lowercase nucleotides are flanking sequences which act as splice junctions).

17.2.4 Modifications to exon-trapping protocol

17.2.4.1 Isolating expressed sequences from YACs

Protocol 91 can also be simply modified to isolate expressed sequences from YACs [36] (see protocol for details). The PCR-amplified material can either be used as an exon-enriched probe [36], or individual products can be subcloned and sequenced. Individual subclones should be hybridized back to a Southern blot containing the tested YAC and a panel of unrelated YACs to confirm their specificity. In addition to the artefacts discussed in Section 17.2.2, some of the clones may correspond to yeast genes (including yeast ribosomal sequences). Because of the yeast background, it is not recommended that yeast genomic DNA containing a YAC of interest is directly subcloned into pSPL1. Exon trapping from YACs is generally less well characterized than trapping from individual cosmids and is technically rather less straightforward. It may, therefore, in many cases be preferable to identify a set of cosmids by hybridization of the YAC Alu-PCR product to a cosmid library. The cosmids can then be trapped in pools of 5–10 or individually.

17.2.4.2 The pSPL3 exon-trapping vector

A modified form of the pSPL1 exon-trapping vector can be obtained from Gibco-BRL. It is referred to as pSPL3 and allows removal of the PCR product generated from nonrecombinant pSPL1 clones or when a particular subcloned genomic fragment does not contain an exon, so that the *tat* splice junctions are paired to generate vector-only product [35]. In pSPL3 a *Bst*XI site is created in the 'vector-only' spliced product but remains as two 'half-sites' in PCR products containing trapped exons. After reverse transcription as described in Protocol 91, six cycles of the primary PCR reaction are completed and vector-only products are cut with *Bst*XI (50 units added directly to the 100 µl PCR reaction and incubated at 55 °C overnight). Ten microlitres of the primary PCR reaction is then added to a 100 µl secondary PCR reaction and cycled 35 times. By removing the vector-only product, the sensitivity of the system is increased considerably, particularly when gene-poor regions are being scanned for exons (when the vast majority of PCR product is vector-only, which can compete out larger products corresponding to trapped exons).

17.3 Direct cDNA hybridization methods

Protocol 92 describes a method for direct cosmid or YAC hybridization to cDNA filters. Given the

potential complications in this strategy outlined in the introduction (cross-hybridization of repeats and low signal-to-noise ratio), it is important to verify that cDNA positives are genuine. The simplest way to check if a cDNA represents an expressed sequence from the region is to use somatic cell hybrids to verify mapping of a clone to the chromosome and region of interest. In the absence of suitable hybrids, single-copy signal on genomic Southernblots together with hybridization to YACs from the region gives a good indication that a cDNA has not been detected by a repetitive cross-hybridization.

17.4 cDNA enrichment

A generally more successful method for direct cDNA isolation from YACs is the solution hybridization approach mentioned previously and described in detail in Protocol 93. Here genomic DNA is biotinylated, hybridized in solution to cDNA, and transcripts hybridizing to genomic DNA are isolated by the addition of streptavidin-coated magnetic beads. The strength and stability of the biotin–streptavidin coupling (the reaction takes place even in 25% formamide) allows DNA manipulations such as washing of heteroduplex DNA at any desired stringency, thermal denaturation and elution of annealed cDNAs, or a simple and efficient change of buffers. Using beads, a very flexible system for blocking repetitive sequences, hybridization, washing and elution is generated. Moreover, the beads used in most of the experiments (Dynabeads M-280, Dynal, Oslo, Norway) are monodispersed and thus follow uniform kinetics when subjected to a magnetic field. It is not necessary to use monodispersed beads in this experiment but the streptavidin beads described above give reproducibly good results.

The genomic DNA used for cDNA selection can be cloned in any genomic cloning system (λ -phages, cosmids, P1 phages or YACs; see, for example, Protocols 78, 85 and 86 in Chapter 15). All genomic sources ultimately result in an enrichment of cDNAs encoded by the insert. Nevertheless, the cDNA enrichment using cosmid or P1 clones as probes is more efficient than with λ -cloned DNA because of the better insert-to-vector ratio. Also, YAC DNA is not as good a starting material as cosmid DNA, since gel-purified YAC DNA always contains degraded yeast DNA from higher molecular weight yeast chromosomes, which contain ribosomal sequences. Therefore, in some of the enriched cDNA sublibraries using YAC DNA as starting material, more than 70% of the clones were of ribosomal origin as a result of the strong homology between yeast

and human ribosomal RNA sequences [14]. These selection artefacts can be overcome in two ways: (i) by counter screening of the sublibrary with ribosomal probes to identify clones containing ribosomal sequences and (ii) by competition of the YAC DNA with total yeast DNA (described in Protocol 92).

Protocol 93 tends to normalize the frequency of the transcripts which are encoded by the genomic source, with greater enrichment factors for rare transcripts than for abundant ones [14]. Enrichment factors of up to 10^5 have already been reported for infrequent cDNA clones and efficiencies as high as 10^6 seem to be within reach. Very little genomic target DNA is needed (positive results can be obtained when using 1 pg cosmid DNA) and increasing genomic target DNA has no or very little effect on coselection of nonspecific cDNA [15]. A reduction in the yield of a specific cDNA (reduced enrichment factor) with more complex DNA targets is due to the competition between the large number of positively selected cDNAs during PCR.

This technique is of most value when the expression pattern of the gene of interest is known, so that the original tissue cDNA library is certain to contain the gene of interest. When the expression pattern of the gene is not known cDNA from different tissues and developmental stages is required. A combination of random-primed, uncloned, double-stranded cDNAs can be used to increase the likelihood of identifying a given gene. A second serious problem inherent in this method is the coselection of pseudogenes. Pseudogenes and other artefacts can be eliminated from subsequent analysis by using exon-trapping and cDNA enrichment in parallel, as described previously.

17.5 Obtaining a full-length cDNA

Obtaining a full-length cDNA is often possible by walking in the enriched library described above or by isolation of large insert cDNA clones from conventional cDNA libraries using exon-trap products, minilibrary cDNA clones, etc. If these relatively straightforward strategies still fail to generate full-length transcripts, PCR-based methods can be used to obtain 5' and 3' cDNA ends. These methods are

referred to as RACE (rapid amplification of cDNA ends).

17.5.1 3' RACE

RNA is reverse transcribed using a primer containing a 3' oligo(dT) stretch and a unique 5' sequence (to increase the specificity of subsequent PCR amplifications). Amplification is subsequently performed using a primer specific to the cDNA sequence and a primer complementary to the unique 5' tail of the oligo(dT) primer (the 'adaptor' primer). A nested PCR reaction can be carried out using a second gene-specific primer 3' to the gene-specific primer used in the primary PCR reaction. A typical reaction scheme is outlined in Protocol 94, although the number of PCR cycles necessary to obtain a product visible on an ethidium-stained gel will vary considerably, depending on the message abundance in the RNA tested. Successful amplification can be performed from total RNA or poly(A)⁺ RNA, and gene-specific primers should be chosen close to the 3' end of the known cDNA sequence and with annealing temperatures similar to that of the adaptor primer.

17.5.2 5' RACE

A modified method for obtaining 5' ends of cDNAs has been described [40]. Here, in Protocol 95, instead of using terminal transferase to add homopolymeric tails to the 3' end of the first-strand cDNA (prior to PCR amplification between a gene-specific primer and a homopolymeric primer complementary to the cDNA tail), a unique oligonucleotide is ligated to the 3' end of the first-strand cDNA using T4 RNA ligase (which is capable of ligating two single-stranded DNA molecules). The method gives significant improvements in the specificity of 5' RACE but requires that the oligonucleotide used is blocked at its 3' end (to prevent concatamerization). In Protocol 95 this is achieved by tailing the oligonucleotide with radioactively labelled dideoxyATP (ddATP). Alternatively, an appropriate oligonucleotide is supplied with the Clontech 5' AmpliFINDER RACE kit.

Protocol 91 Exon trapping by pSPL1 from genomic DNA cloned in cosmids

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Overview

- (a) Construction of pSPL1 recombinants
- (b) Transient expression of pSPL1 library in mammalian cells and RNA extraction
- (c) cDNA synthesis from cytoplasmic RNA
- (d) Cloning of exon DNA
- (e) Modifications for isolating expressed sequences from YACs

(a) Construction of pSPL1 recombinants

Materials

- cos mid DNA
- plasmid pSPL1 (Gibco-BRL)
- restriction enzymes *Bam*HI and *Bg*II
- restriction enzyme buffer
- Strataclean (Stratagene)
- phenol/chloroform
- chloroform
- ethanol
- calf intestinal alkaline phosphatase (CIP)
- T4 DNA ligase (5 U μ l⁻¹)
- 5 \times ligation buffer: 250 mM Tris-HCl (pH 7.6), 50 mM MgCl₂, 5 mM rATP, 5 mM DTT, 25% PEG 8000
- equipment for electroporation (e.g. BioRad Gene Pulser)
- equipment for agarose gel electrophoresis
- *E. coli* DH5 α cells
- Qiagen-20 column (Qiagen)
- 1 \times TE: 10 mM Tris-HCl (pH 7.5), 1 mM EDTA

Method

LIGATION OF COSMID DNA INTO pSPL1

- 1** Digest about 1 μ g cosmid DNA with *Bam*HI and *Bg*II in a final volume of 100 μ l following manufacturer's protocols (or see ref. 1 for details of procedures for restriction digestion).
- 2** Strataclean the digest twice following the manufacturer's protocol. Alternatively, remove restriction enzyme with phenol/chloroform,

chloroform and ethanol precipitation using standard methods. Check the success of the digest by running an aliquot on an agarose gel.

3 Cut the pSPL1 vector with *Bam*HI and dephosphorylate the 5' ends with calf intestinal phosphatase (CIP) following standard methods.

4 Ligate as follows:

- 10 µl cosmid digest (100 ng);
- 1 µl pSPL1 (50 ng): phosphatased;
- 4 µl 5× ligation buffer;
- 4 µl water;
- 1 µl T4 DNA ligase (5 U µl⁻¹);

20 µl total.

Ligate for 3 h at room temperature and in parallel run the following controls:

- 1 no cosmid DNA;
- 2 no pSPL1;
- 3 no ligase.

TRANSFORMATION OF *ESCHERICHIA COLI* WITH RECOMBINANT pSPL1

5 Electroporate 1 µl of the ligation into 40 µl of electrocompetent *E. coli* DH5α cells and shake for 1 h in 1 ml LB broth. Spin to collect cells, resuspend in 100 µl, spread on an LB plate supplemented with ampicillin (100 µg ml⁻¹) and allow to grow overnight at 37 °C. Electroporation conditions (using a BioRad Gene Pulser and an electroporation cuvette with a 0.2-cm gap width) are as follows:

- 2.5 kV;
- 25 µF;
- 200 ohm;
- about 4.7 s time constant.

Other transformation procedures may also be used.

6 Scrape colonies from the plate and purify the pSPL1 library DNA using a Qiagen-20 column following the manufacturer's recommendations. Other alkaline lysis extraction methods may also be used (see, e.g. Chapter 15, Protocol 87; Chapter 21, Protocol 100). Resuspend the purified DNA in 1×TE at 100 ng µl⁻¹.

You should get a few clones (< 10) on the pSPL1-only plate and the cosmid-only plate and (an absolute minimum of) more than 100 clones on the cosmid/pSPL1 ligation plate. The more recombinants the better, as this increases the probability of including large genomic fragments in the pSPL1 sublibrary (large fragments are less efficiently subcloned than small fragments, and insert sizes of greater than 5 kb are rarely observed).

When initially establishing this method with a particular batch of phosphatased vector, it is helpful to test individual clones by digestion with *Sal*II/*Nde*I. Digest pSPL1 plasmid DNA as well as a control. Run the samples on a 1% agarose gel and expect the 600-bp band (in non-recombinants) to be shifted in size in recombinant clones. Most (> 90%) clones should be recombinants with a variety of insert sizes.

(b) Transient expression of pSPL1 library in mammalian cells and RNA extraction

The pSPL1 library is now ready for electroporation into mammalian cells.

Materials

- COS-7 cells (available from the American Type Culture Collection, Rockville, MD; see Appendix V for address)
- Dulbecco's Modified Eagle's Medium (DMEM)
- 0.25% trypsin
- 1 M EDTA
- PBSA
- PBS
- TKM
- Triton X-100
- Tris-buffered phenol
- 5% SDS
- phenol/chloroform/isoamylalcohol (25:24:1)
- 5 M NaCl
- absolute ethanol
- RNase-free water
- Eppendorf tubes
- centrifuge
- equipment for electroporation (e.g. BioRad Gene Pulser)

Method

ELECTROPORATION OF pSPL1 DNA INTO MAMMALIAN CELLS

- 1 COS-7 cells are grown in Dulbecco's Modified Eagle's Medium (DMEM) supplemented as described. Cells should be grown at 37°C in a 5% CO₂ incubator. For each electroporation 1–5 × 10⁶ cells from a 60–80% confluent culture are required. For example, a 160-cm² tissue culture flask at appropriate density is sufficient for three electroporations. General methods for passaging and maintaining mammalian cell lines are described in ref. 32.
- 2 Medium is removed from COS-7 cells and replaced with 5 ml 0.25% (w/v) trypsin, 1 mM EDTA. Detach the layer of cells by agitation and then add 5 ml of DMEM. Pellet the cell suspension (300 g, 5 min) and resuspend in 40 ml ice-cold PBSA. Estimate cell density, spin as above and resuspend in ice-cold PBSA at 1–5 × 10⁶ cells ml⁻¹. Mix 1 ml of cells with the DNA to be transfected (1–20 µg brought to 50 µl with PBSA), transfer to a 0.4-cm electroporation cuvette (prechilled) and electroporate (BioRad Gene Pulser) using the conditions below:
 - 1.2 kV;
 - 25 µF;

- 200 ohm;
- time constant 0.9–1 s.

After electroporation, keep the cells on ice for 10 min then transfer to a small tissue culture dish containing 10 ml prewarmed, pre-equilibrated DMEM. Incubate for 2 days, after which the cells should be $\approx 50\%$ confluent.

CYTOPLASMIC TOTAL RNA EXTRACTION

- 3 Remove supplemented DMEM from each plate and wash three times with ice-cold PBS. Place plates on a bed of ice, add 10 ml PBSA and scrape cells into a suspension. Transfer cells to a 15-ml conical centrifuge tube and spin for 5 min, 300 g, 4 °C.
- 4 Place tubes on ice and decant supernatant. Resuspend cells in 300 μ l TKM and incubate on ice for 5 min.
- 5 Add 15 μ l of 10% (v/v) Triton X-100, mix and incubate on ice for an additional 5 min.
- 6 Centrifuge for 5 min at 450 g to pellet the nuclei and transfer the supernatant to an Eppendorf tube (on ice) containing 20 μ l of 5% SDS, 300 μ l of Tris-buffered phenol, vortex mix and spin in a microcentrifuge for 5 min.
- 7 Transfer the supernatant to a second (ice-cold) tube containing 300 μ l phenol/chloroform/isoamylalcohol (25 : 24 : 1), vortex mix and separate the phases by centrifugation as above.
- 8 Transfer the upper aqueous layer to a third Eppendorf tube containing 12 μ l 5 M NaCl, add 750 μ l absolute ethanol and mix well by vortexing. Incubate on dry ice for 10 min, spin at top speed in a microcentrifuge (13 000 g) for 15 min.
- 9 Dry pellet and resuspend in 20 μ l of RNase-free water. Store at -70°C .

Other methods for obtaining cytoplasmic RNA can also be used (see, for example, Chapter 18, Protocol 96).

(c) cDNA synthesis from cytoplasmic RNA

This reverse transcription/PCR amplification procedure (RT-PCR) is robust: DNA contamination is not important and the amplification is successful even with fairly degraded RNA.

Materials

- cytoplasmic RNA as prepared in (b)
- 5 \times reverse transcription buffer: 250 mM Tris-HCl (pH 8.3), 400 mM KCl, 15 mM magnesium chloride, 50 mM DTT
- 10 \times PCR buffer: 500 mM KCl, 100 mM Tris-HCl (pH 9.0) at 25°C, 15 mM MgCl_2 , 1.0% Triton X-100

- DTT
- pSPL1 oligonucleotides:
 - SA4: 5' CACCTGAGGAGTGAATTGGTCG 3'
 - SD2: 5' GTGAACTGCACTGTGACAAGCTGC 3'
 - SD1: 5' CUACUACUACUAGCGACGAAGACCTCCTCAAGGC 3'
 - SA1: 5' CAUCAUCAUGTCGGGTCCCTCGGGATTGG 3'
- dNTPs
- RNAsin (BRL)
- Moloney murine leukaemia virus (MMLV) reverse transcriptase
- *Taq* DNA polymerase
- 5×TBE: 54 g l⁻¹ Tris base, 27.5 g l⁻¹ boric acid, 20 ml l⁻¹ EDTA (0.5 M) (pH 8.0)
- 1.5% low-melting-point agarose gel (SeaPlaque CTG)
- GeneClean 'glass milk' (BIO 101)

Method

REVERSE TRANSCRIPTION OF CYTOPLASMIC RNA

- 1** Reverse transcription: Add the following components into a microcentrifuge tube:
 - 2.5 µl 10× reverse transcription buffer;
 - 1 µl cytoplasmic RNA as prepared above (b);
 - 1 µl DTT (100 mM);
 - 1.25 µl oligonucleotide SA4 (20 µM);
 - 2 µl dNTPs (2.5 mM);
 - 15.25 µl distilled water;
 23 µl total.
- 2** Denature the RNA by heating the sample to 65 °C for about 3 min. Allow samples to cool to room temperature then add: 1 µl RNAsin and 1 µl Moloney murine leukaemia virus (MMLV) reverse transcriptase. Mix by vortexing briefly and incubate at 42 °C for 90 min.

PCR AMPLIFICATION OF REVERSE-TRANSCRIBED EXON DNA

- 3** Primary PCR reaction:
 - 25 µl reverse transcription reaction;
 - 7.5 µl 10× PCR buffer;
 - 6 µl dNTPs (2.5 mM);
 - 5 µl oligonucleotide SD2 (20 µM);
 - 3.75 µl oligonucleotide SA4 (20 µM);
 - 52.25 µl distilled water;
 - 0.5 µl *Taq* DNA polymerase (5 U µl⁻¹);
 100 µl total.
- Cycle 35 times as follows using a thermal cycler:
- 94 °C for 1 min;
 - 58 °C for 1 min;
 - 72 °C for 2 min.

4 Secondary PCR reaction: transfer 1 μ l of the above reaction directly into the following 50 μ l PCR reaction mix:

- 5 μ l 10 \times PCR buffer;
- 4 μ l dNTPs (2.5 mM);
- 2 μ l oligonucleotide SA1 (20 μ M);
- 2 μ l oligonucleotide SD1 (20 μ M);
- 36.5 μ l distilled water;
- 0.5 μ l *Taq* DNA polymerase (5 U μ l⁻¹);

This 'nested' PCR reaction is cycled 10–15 times with the above conditions.

5 Run 10- μ l samples on a 1.5% low-melting-point agarose gel in 1 \times TBE buffer with the inclusion of appropriate size markers (e.g. the 100-bp ladder from Gibco-BRL). Fragments of size greater than 130 bp (the vector-only splice) are candidate 'trapped' exons and may be rapidly gel-purified using GeneClean 'glass milk'. Examples of exon-trapped products are shown in Fig.17.3. These fragments can be rapidly subcloned as the primers used in the secondary PCR reaction contain dUMP residues [33].

(d) Cloning of exon DNA

After treatment with uracil DNA glycosylase (UDG), compatible single-stranded ends are generated between the exon amplification product and the UDG cloning vector pAMP.

Materials

- Genecleaned secondary PCR product
- pAMP cloning vector (Gibco-BRL)
- 10 \times PCR buffer
- uracil DNA glycosylase (UDG)
- *E. coli* DH5 α cells
- LB agarose containing 100 μ g ml⁻¹ ampicillin

Method

1 Mix 5 μ l (10–100 ng) Genecleaned secondary PCR product, 1 μ l pAMP cloning vector (50 ng μ l⁻¹), 1 μ l 10 \times PCR buffer, 2 μ l distilled water, 1 μ l UDG (1 U μ l⁻¹) and incubate at 37 °C for 45 min. Cool on ice and transform 1–5 μ l into *E. coli* DH5 α cells and plate on LB agarose containing 100 μ g ml⁻¹ ampicillin. Essentially all ampicillin-resistant colonies are recombinants and can be further characterized by sequencing and hybridization to cDNA libraries.

Hybridization to Northern blots and zoo blots can also be attempted but the small size of the amplified exons can lead to sporadic failure and is not recommended [6]. Direct hybridization to cDNA libraries or, preferably, enriched minilibraries (see Section 17.1.3 and Protocol 93) is

technically much more straightforward. As most exons are relatively small (< 200 bp), they can be rapidly sequenced in large numbers and nucleic acid or protein database searches can lead to the identification of related genes of known function.

Extensive characterization of the system indicates that virtually all exon amplification products which do not fall into either of the two common artefact categories described in Section 17.2.2 are genuine expressed sequences.

(e) Modifications for isolating expressed sequences from YACs

Materials

- low-melting-point agar (SeaPlaque CTG)
- restriction enzyme digestion buffer
- *Bam*HI
- *Bg*II
- Geneclean (BIO 101)
- phosphatased pSPL1 vector

Method

- 1 Isolate the YAC in low-melting-point agarose by preparative PFGE and cut an agarose gel slice containing the YAC under long-wave UV. Ideally, conditions should be chosen under which the YAC is clearly separated from host chromosomes.
- 2 Incubate the gel slice in restriction enzyme digestion buffer for 2 h then digest with *Bam*HI/*Bg*II, removing the gel slice into a new tube containing restriction buffer and enzymes.
- 3 Incubate for 8 h and then isolate the digested DNA using Geneclean. Elute the DNA in 50 µl water according to the manufacturer's instructions and ligate 10 µl into the phosphatased pSPL1 vector following standard methods (Protocol 70a). Obtaining sufficient clone numbers for at least a fivefold coverage of the YAC may require scaled-up ligation volumes and precipitation before electroporation.

Once a sufficiently representative library has been generated, the method follows exactly the same steps as detailed for single cosmids.

.....

Troubleshooting

Ratio of recombinants to non-recombinants is less than 10 : 1

- Repeat the pSPL1-phosphatasing step with a new batch of phosphatase and repeat the ligation with new ligase and ligase buffer.

If low recombinant to non-recombinant ratios persist, ensure that the vector ends are intact by treating with polynucleotide kinase and religating a sample, and ensure that the restriction enzymes used in this procedure have been thoroughly removed before ligation (Strataclean, ethanol precipitate, and ligate at 14 °C overnight).

.....

Protocol 92 Direct cosmid or YAC hybridization to cDNA filters

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- cDNA library on filters
- YAC or cosmid DNA in agarose blocks (see Chapter 15, Protocols 81 and 85)
- Church buffer: 0.5 M sodium phosphate (pH 7.2), 7% SDS, 1 mM EDTA (pH 8.0)
- UV light source
- ethidium bromide
- Geneclean (BIO 101)
- [³²P]dCTP and [³²P]dATP
- TE (see Protocol 91)
- sonicated genomic DNA (≈ 300 bp)
- Cot-1 DNA (Gibco-BRL)
- yeast tRNA
- cDNA library vector DNA (digested or sheared)
- 1 M sodium phosphate buffer, pH 7.2
- 0.1% SDS
- SSC buffers
- equipment for autoradiography

Method

- 1 Half a million primary or 1 million amplified cDNA clones should be screened at a rate of 200 000 clones per 22 × 22 cm membrane. Duplicate filters are required.
- 2 Isolate the YAC on a low-melting-point pulsed-field gel (10 blocks side by side in a single slot). Use conditions that provide an optimal separation of the YAC from yeast chromosomes. Best results will be obtained with good quality, undegraded agarose YAC blocks, each containing 2 ml of saturated yeast culture per block.
- 3 Visualize the YAC band using long-wave UV light after staining the gel with ethidium bromide and destaining with water, and excise

the band in a minimal volume of agarose. If there is difficulty in identifying the YAC band, cut a strip off the gel and view it under short wave UV light. Use this strip to mark the position of the YAC band so that it can be excised without visualization. Check that the YAC has been excised by photographing the remaining gel under short-wave UV light. Size markers and extra tracks carrying the YAC of interest and another, different, YAC will help to ensure that no mistakes are made.

- 4 GeneClean (BIO 101) the YAC DNA from the agarose in order to concentrate and shear it.
- 5 Take 25% of the DNA and radiolabel by random priming [37] overnight using [³²P]dCTP and [³²P]dATP.
- 6 Purify the radiolabelled probe by precipitation and resuspend in 100 µl of 1×TE. Specific activity should be greater than 10⁹ c.p.m. µg⁻¹.
- 7 Prehybridize cDNA library filters in Church buffer in plastic bags. A maximum of 10 filters per plastic bag ensures good distribution of the hybridization fluid (20–30 ml) throughout the membranes. Sandwich the bags between glass plates and gently rock (rocking is not essential) at 65 °C.
- 8 Preanneal the probe to suppress repeat and vector hybridization [38,39]. Add and mix the following to 100 µl of probe:
100 µg sonicated genomic DNA (average size, 300 bp), 100 µg Cot-1 DNA (Gibco-BRL), 100 µg yeast tRNA, 10 µg cDNA library vector DNA (digested or sheared). Heat at 100 °C for 10 min.

Add sodium phosphate buffer (1 M stock, pH 7.2) to give a final concentration of 0.12 M. The total volume should be around 300 µl. Incubate at 65 °C for 2 h.
- 9 Add preannealed probe to 20–30 ml of Church buffer. Remove prehybridization buffer from plastic bags containing cDNA filters and replace with probe buffer. Ensure that the probe is thoroughly distributed throughout the membranes. Hybridize overnight at 65 °C.
- 10 Wash filters in 40 mM sodium phosphate (pH 7.2), 0.1% SDS at room temperature for 2×15 min followed by 15 min at 65 °C. Expose on preflashed fast film (Kodak X-AR) for 1–5 days. Membranes can be rewashed at higher temperatures or in higher stringency SSC buffers (40 mM phosphate = 0.2×SSC) after initial autoradiography.

.....

Protocol 93 cDNA enrichment using solution hybridization to genomic clones as probes

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Overview

- (a) Preparation of probe
- (b) Preparation of cDNA
- (c) Enrichment of cDNA by solution hybridization to genomic probe

(a) Preparation of probe

Materials

- DNA (cosmid, cosmid pool or YAC)
- 10× nick translation buffer (NTB): 500 mM Tris-HCl (pH 7.5), 50 mM MgCl₂, 500 µg ml⁻¹ BSA
- 10× nucleotide mix with biotin (NTB-BIO): 500 µM dCTP, 500 µM dGTP, 500 µM dATP, 380 µM dTTP, 30 µM biotin-16-dUTP (Amersham)
- 5× ligation buffer (see Protocol 91)
- DTT
- DNase
- *E. coli* DNA polymerase I
- EDTA, pH 8.0
- 3 M sodium acetate, pH 4.8
- ethanol
- mRNA
- hexamers
- 5× reverse transcription buffer (see Protocol 91)
- reverse transcriptase (Superscript, BRL)
- 10 mM dNTPs
- 25 mM Tris-HCl (pH 7.5)
- 100 mM KCl
- 5 mM MgCl₂
- 150 µM β-NAD⁺
- 10 mM ammonium sulphate
- RNase H
- *E. coli* DNA ligase
- T4 DNA polymerase
- phenol/chloroform
- 7.5 M ammonium acetate
- T4 DNA ligase
- adaptors of choice
- 0.75 M NaCl
- 50 mM sodium phosphate, pH 7.2
- 5× Denhardt's solution
- 50% formamide

- streptavidin beads (Dyna)
- B&W very high salt buffer (Dyna)
- biotin
- magnetic device for collecting beads (MPC, Dyna)
- $2\times\text{SSC}$
- $0.2\times\text{SSC}$
- $0.1\times\text{SSC}$
- materials for PCR
- CloneAmp system (BRL)
- Eppendorf centrifuge

Method

- 1 Biotinylate genomic clones by nick translation. Photobiotinylation also works well, but is very vulnerable to contamination with proteins and amino groups in Tris buffers. In general, cosmids work best (using a single cosmid, an enrichment as high as 10^5 has been achieved; using cosmid pools spanning 500 kb of genomic DNA the enrichment is in the range of 5×10^3). YACs give considerably higher background of nonspecific clones and require modifications described later. An even incorporation of biotin molecules across the genomic clones is desirable, to bind longer genomic fragments to the same extent as shorter ones. In our hands an incorporation frequency of one biotin per 100 bp gives reproducibly good results and does not interfere with the hybridization of nucleic acids.
 Mix on ice: 1 μg DNA (cosmid, cosmid pool or YAC), 5 μl $10\times\text{NTB}$, 5 μl $10\times\text{NTB-BIO}$, 5 μl 0.1 M DTT, 1–20 μl DNase (1 ng μl^{-1}), 2 μl *E. coli* DNA polymerase I (10 U μl^{-1}).
 Add water to 50 μl final volume and incubate at 15 °C for 3 h and stop the reaction by the addition of 5 μl 0.5 M EDTA (pH 8.0). Check 15 μl on gel to determine the size of the fragments. The amount of DNase required must be determined experimentally to generate an average fragment size of 1 kb. Precipitate with 4 μl 3 M sodium acetate (pH 4.8) and 120 μl ethanol. Incubate at –80 °C for at least 10 min and spin at top speed in a benchtop Eppendorf centrifuge at 4 °C, for at least 20 min. Wash pellet with 70% ethanol and air-dry for 10 min, store dry at –20 °C.

(b) Preparation of cDNA

- 2 The cDNA source (cDNA amplified with adaptor or vector primers) is an important component in determining the success of the experiment. cDNA is prepared by directly reverse transcribing an appropriate mRNA source and is amplified without a cloning step (by ligating adaptors to the double-stranded cDNA molecules). Use high temperature annealing primers (> 60 °C), and preferentially a two-step PCR (72 °C and 94 °C cycling) to avoid nonspecific amplification. Whenever possible, use randomly primed cDNA as

cDNA amplification products longer than 1.5–2 kb are rarely obtained. Also, the use of oligo(dT)-primed cDNA (libraries), results in sublibraries enriched for clones from 3' untranslated regions of genes, which are generally less informative than translated sequences. Nevertheless, any source of cDNA (uncloned or cloned) can be used in a selection experiment, under the condition that it is PCR-amplifiable. Inserts from an oligo(dT)-primed cDNA library cloned in gt10 [13], a normalized short insert library [15], *Mbo*I-digested, oligo(dT)-primed cDNA [14], or a combination of random and oligo(dT)-primed human fetal brain library [17] have been reported thus far. In addition, cDNAs from different sources cloned in gt10, gt11, SWAJ-2, -ZAP, and pCDM8 have been used with reproducibly good results.

Typically, as described in detail below, random-primed cDNA is ligated to adaptors and PCR amplified, giving average insert sizes of about 750 bp and single clones as long as 2.4 kb.

- 3** First strand reverse transcription reaction: 2.5 µg mRNA, 2 µg hexamers (100 ng µl⁻¹), 4 µl 5× reverse transcription buffer, 2 µl dNTPs (10 mM), 2 µl reverse transcriptase.
Add water to 20 µl final volume and incubate at 37 °C for 1 h. Stop the reaction by placing on wet ice.
- 4** For second-strand synthesis, dilute the first-strand reaction to 150 µl with the final composition of: 25 mM Tris-HCl (pH 7.5), 100 mM KCL, 5 mM MgCl₂, 1.2 mM DTT, 250 µM each dNTP, 150 µM β-NAD⁺, 10 mM ammonium sulphate and add 2 U RNase H, 10 U *E. coli* DNA ligase, 40 U *E. coli* DNA polymerase.
- 5** Incubate for 2 h at 15 °C, then add 10 U T4 DNA polymerase and incubate for 5 more minutes. The reaction is stopped with 5 µl 0.5 M EDTA (pH 8.0), extracted with phenol/chloroform and precipitated with 0.5 vols 7.5 M ammonium acetate and 3 vols of ethanol. After washing with 70% ethanol the pellet is dried.
- 6** The dry double-stranded cDNA is ligated to an adaptor of choice. The adaptor should carry a sequence allowing it to prime in a PCR reaction at a temperature above 60 °C. Take up the cDNA in a 50-µl solution containing 1 nmol of adaptor, and add: 11 µl 5× ligation buffer and 5 µl T4 DNA ligase (highest concentration available).
Incubate for 16 h at 16 °C, then extract once with phenol/chloroform and precipitate with 0.5 vols 7.5 M ammonium acetate and 3 vols ethanol. After washing with 70% ethanol, air dry the pellet and resuspend in 20 µl 1× TE.
- 7** Amplification of ligated cDNA is carried out using 0.5 µl of the ligation product per 100 µl PCR reaction and 1 µg of each adaptor primer. Amplify for 25 cycles with an extension time of at least 3 min. Check one aliquot of the product and precipitate in 10 µg aliquots, leave the cDNA under ethanol at -20 °C until it is to be used.

(c) **Enrichment of cDNA by solution hybridization to genomic probe**

In a typical experiment use 0.2–1 ng genomic DNA per kb (e.g. for a cDNA fishing experiment using a single 40 kb cosmid, use 8–40 ng of the biotinylated product). Repetitive sequences must first be removed, and competition of the genomic sample is carried out with human Cot-1 DNA (250- to 500-fold excess by mass), vector DNA (100-fold excess by mass), and yeast DNA (100-fold excess, when using a YAC). Prepare enough cDNA for two rounds of the enrichment (25-fold excess by mass over genomic DNA). While the method described here is very effective, it is still unable to suppress all low copy repeats or repeats which are only present in certain subchromosomal regions [16]. The competition volume is 25–75 μ l, depending on the amount of DNA included, but always have a final concentration of 2–4 μ g DNA per microlitre. The hybridization solution is: 0.75 M NaCl, 50 mM NaPO₄ (pH 7.2), 0.05% SDS, 5 \times Denhardt's, 1 mM EDTA (pH 8.0), 50% formamide.

- 8** Denature the competition DNA together with the biotinylated cosmids in the appropriate amount of hybridization solution for 10 min at 80 °C, cool on ice, spin down and hybridize for 2 h at 42 °C.
- 9** Wash an appropriate amount of streptavidin beads (5–50 μ l) three times in 500 ml B&W to remove preservative and then resuspend in B&W. Add the washed beads (3–5 vols) to the hybridization solution and bind biotin at room temperature for at least 45 min. Collect the beads using a magnetic device for reaction tubes and remove liquid.
- 10** Precipitate an appropriate amount of cDNA, at least 50-fold excess by mass over genomic DNA, and resuspend in 15–50 μ l hybridization solution with a cDNA concentration of at least 1 mg ml⁻¹. Denature at 80 °C for 10 min, cool on ice for 1 min, spin down briefly and resuspend the magnetic beads in the solution. Incubate for at least 16 h at 42 °C.
- 11** After overnight hybridization, wash beads with 2 \times SSC, 0.2 \times SSC and 0.1 \times SSC (500 μ l each) for 10 min at 65–68 °C. Repeat the high stringency wash at least four times and elute specific cDNAs in water (50–100 μ l) at 85 °C for 10 min.
- 12** Amplify 5 μ l of the elution per 100 μ l PCR reaction using appropriate primers and conditions for 20–25 cycles with an elongation time of at least 3 min.
- 13** Check PCR products on gel for size and estimate concentration.
- 14** Run a second round of cDNA enrichment using fresh genomic DNA and fresh beads. After the second round you should see a clear banding pattern in the amplified cDNA, when starting from a cloned cDNA library. Starting with uncloned cDNA will give a smear between 0.1 and 2.5 kb. Optionally, a third round of selection/amplification can be done, but this does not always result in a further enrichment.

- 15** Finally, clone the amplified cDNA (using the BRL CloneAmp system). Before this step it is advantageous to do a size selection of the PCR products either on a gel or on a sizing column (e.g. Sephacryl S-400, Pharmacia).
- 16** Pick recombinants into microtitre plates taking at least 1.5–2 cDNA clones per kb of genomic DNA used. Filters can be prepared by robotically or manually replicating clones from microtitre plates onto hybridization membranes. It is important to do a control screen of these enriched cDNA library filters with a human Cot-1 probe to detect repetitive clones (which should not represent more than 15–20% of the cDNAs). Do a second control screen using the cloning vector for the genomic DNA source. Also, when cDNA selection is attempted with a YAC, screening with a ribosomal DNA probe is helpful as up to 40% of the clones can be of ribosomal DNA origin. Typically, expect 60–80% of the cDNA to represent genuine transcripts which map back to the genomic region tested in the experiment. Lower figures are obtained with YACs.

.....

Protocol 94 Rapid amplification of 3' ends

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Overview

- (a)** Reverse transcription
- (b)** Primary PCR reaction

(a) Reverse transcription

Materials

- 5×reverse transcription buffer (see Protocol 91)
- oligo(dT) adaptor primer: 5' CAGTCAAGGACGCTCATTCGA(T)₁₇ 3'
- total RNA or poly(A)⁺ fraction
- distilled water (RNase-free)
- RNAsin (BRL)
- Moloney murine leukaemia virus reverse transcriptase (BRL)
- TE (see Protocol 91)

Method

- 1** Mix 4 µl 5×reverse transcription buffer, 1 µl oligo(dT) adaptor primer (100 ng µl⁻¹), 2 µl total RNA (1 µg µl⁻¹) or 100 ng of a poly(A)⁺ fraction, 10 µl water (RNase-free).

Heat to 65 °C for 5 min and cool on wet ice for 1 min. Add 1 µl RNAsin and 2 µl Moloney murine leukaemia virus reverse transcriptase.

- 2 Incubate at 42 °C for 1 h and add 500 µl 1×TE. Store at –20 °C. A single reverse transcription generates enough cDNA for many RACE reactions. Include the following controls: (a) no RNA and (b) replacement of RNA with genomic DNA. If either of these samples gives a visible product after the PCR reactions described below, the experiment is compromised, either because the primers are not separated by an intron in genomic DNA or because they generate nonspecific amplification products. New amplification primers should be designed.

(b) Primary PCR reaction

Materials

- reverse transcription reaction (obtained in steps 1–2 of (a))
- 10×PCR buffer (see Protocol 91c)
- dNTPs (2.5 mM)
- 5' gene-specific primer (20 µM)
- 3' gene-specific primer (20 µM)
- adaptor primer (20 µM): 5' CAGTCAAGGACGCTCATTCGA 3'
- *Taq* DNA polymerase (5 U µl⁻¹)

Method

- 1 Mix 1 µl reverse transcription reaction, 5 µl 10×PCR buffer, 4 µl dNTPs, 2 µl 5' gene-specific primer, 2 µl adaptor primer, 36.5 µl distilled water, 0.5 µl *Taq* DNA polymerase.
- 2 Amplify for 30–40 cycles with the following conditions:
- 94°C for 1 min;
 - 55°C for 1 min;
 - 72°C for 3 min.
- 3 For a nested PCR use 1 µl of the primary PCR reaction in a 100-µl amplification mix containing the adaptor primer and a gene-specific primer 3' to the primer used in the primary reaction. Titrate the number of cycles (typically, 10–25) required to obtain a visible product with minimal background.
- 4 Products can be gel purified and cloned by the methods described in Protocol 93 and tested for authenticity by sequence comparison to the previously identified cDNA sequence. More than one size of product can be obtained as a result either of multiple polyadenylation signals or of the internal priming of the oligo(dT) to A-rich regions of the message upstream of the poly(A) tail. Sequencing is often the best way to differentiate between these possibilities.

Protocol 95 5' RACE using a unique oligonucleotide ligated to the 3' end of the first-strand cDNA

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Overview

- (a) Reverse transcription
- (b) ddATP tailing of primer
- (c) Single-stranded ligation and PCR amplification

(a) Reverse transcription

Materials

- materials for reverse transcription as in Protocol 94, steps 1–2 but with the oligo(dT) primer replaced with a primer in an antisense orientation to the cDNA coding sequence (50–100 bp from the 5' end).
- poly(A)⁺ RNA
- 5 M NaOH
- 5 M acetic acid
- TE (see Protocol 91)
- centricon spin filter (Amicon)
- glycogen carrier

Method

- 1 Reverse transcription is carried out as described in Protocol 94, steps 1–2, replacing the oligo(dT) primer with a primer in an antisense orientation to the cDNA coding sequence (50–100 bp from the 5' end), and 1–2 µg of poly(A)⁺ RNA is used in preference to total RNA. Use ≈ 1 µl of a 10 µM solution of the antisense primer in the reverse transcription reaction.
- 2 RNA is then hydrolysed by adding 5 µl of 5 M NaOH and incubating at 65 °C for 30 min. The reaction is neutralized by the addition of 5 µl 5 M acetic acid. The sample is diluted to 500 µl with 1 × TE and excess primer removed using a centricon spin filter at 1000 *g* for 20 min. Other forms of column purification (e.g. Sephacryl, Pharmacia) or 'glassmilk' purification (e.g. Geno-Bind, Clontech) may also be used to remove excess primer. The cDNA is recovered by precipitation, including a glycogen carrier (15 µg) and resuspended in 10 µl 1 × TE.

(b) ddATP tailing of primer

Materials

- primer P1: 5'
GCATTGCATCATGATCGATCGAATTCTTTAGTGAGGGTTAATTGCC 3',

- with a 5' phosphate
- terminal transferase (BRL)
- 5×tailing buffer (BRL)
- 500 mM ddATP
- [α - 32 P]ddATP (Amersham)
- TE (see Protocol 91)

Method

- 3** Treat 500 ng of P1 with terminal transferase in a 20 μ l reaction with 4 μ l of 5×tailing buffer, 4 μ l 500 mM ddATP, 1 μ l [α - 32 P]ddATP and 1 μ l terminal transferase. Incubate for 1 h at 37 °C, heat to 75 °C for 10 min and gel purify the labelled oligonucleotide using standard methods [36]. Resuspend at 10 pmol μ l⁻¹ in 1×TE and store at -20 °C.

(c) **Single-stranded ligation and PCR amplification**

Materials

- cDNA (prepared as in a above)
- radiolabelled P1 (prepared as in b above) (10 pmol μ l⁻¹)
- 2×single-stranded ligation buffer
- T4 RNA ligase (10 U)
- primers for PCR, e.g.:
P2: 5' GGCAATTAACCTCACTAAAG 3'
P3: 5' TCACTAAAGAATTCGATCGATC 3'
P4: 5' CGATCGATCATGATGCAATGC 3'

Method

- 4** Mix 3 μ l cDNA, 1 μ l radiolabelled P1, 5 μ l 2×single-stranded ligation buffer and 1 μ l T4 RNA ligase. Incubate at room temperature overnight.
- 5** A PCR reaction is carried out with an antisense gene-specific primer 5' to the primer used in the reverse transcription and a primer of equivalent annealing temperature chosen to be the reverse complement of the P1 sequence. Typically, use 1 μ l of the single-stranded ligation in a 50 μ l PCR reaction with 2 μ l of 20 μ M primer and cycle 35–40 times. Primers P2, P3 and P4 listed above are examples of primers successfully used within the P1 sequence.
- 6** Further nested PCR reactions (using primers P2–P4 listed above) may be necessary to obtain sufficient product from rare transcripts. Specific amplification products can be subcloned and sequenced.

.....

References

- Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Press, Cold Spring Harbor, NY).
- Vulpe, C., Levinson, B., Whitney, S. *et al.* (1993) Isolation of a candidate gene for Menkes Disease and evidence that it encodes a copper-transporting ATPase. *Nature Genet.* **3**, 7–13.
- Trofatter, J.A., MacCollin, M.M., Rutter, J.L. *et al.* (1993) A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumour suppressor. *Cell* **72**, 791–800.
- The Huntington's Disease Collaborative Research Group. (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983.
- Auch, D. & Reth, M. (1991) Exon trap cloning: using PCR to rapidly detect and clone exons from genomic DNA fragments. *Nucleic Acids Res.* **18**, 6743–6744.
- Hamaguchi, M., Sakamoto, H., Tsuruta, H. *et al.* (1992) Establishment of a highly sensitive and specific exon-trapping system. *Proc. Natl Acad. Sci. USA* **89**, 9779–9783.
- Buckler, A.J., Chang, D.D., Graw, S.L. *et al.* (1991) Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl Acad. Sci. USA* **88**, 4005–4009.
- Duyk, G.M., Kim, S., Myers, R.M. & Cox, D.R. (1990) Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. Natl Acad. Sci. USA* **87**, 8995–8999.
- Kelly, A., Powis, S.H., Glynne, R. *et al.* (1991) Second proteasome related gene in the human MHC class II region. *Nature* **353**, 667–668.
- Glynne, R., Powis, S.H., Beck, S. *et al.* (1991) A proteasome-related gene between the two ABC transporter loci in the class II region of the human MHC. *Nature* **353**, 357–360.
- Trowsdale, J., Hanson, I., Mockridge, I. *et al.* (1990) Sequences encoded in the class II region of the MHC related to the ABC superfamily transporters. *Nature* **348**, 741–744.
- Elvin, P., Slyn, G., Black, D. *et al.* (1990) Isolation of cDNA clones using yeast artificial chromosome probes. *Nucleic Acids Res.* **18**, 3913–3917.
- Lovett, M., Kere, J. & Hinton, L.M. (1991) Direct selection: a method for the selection of cDNAs encoded by large genomic regions. *Proc. Natl Acad. Sci. USA* **88**, 9628–9632.
- Morgan, J.G., Dolganov, G.M., Robbins, S.E. *et al.* (1992) The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. *Nucleic Acids Res.* **20**, 5173–5179.
- Parimoo, S., Patanjali, S.R., Shukla, H. *et al.* (1991) cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl Acad. Sci. USA* **88**, 9623–9627.
- Tagle, D.A., Swaroop, M., Lovett, M. & Collins, F. (1993) Magnetic bead capture of expressed sequences encoded within large genomic segments. *Nature* **361**, 751–753.
- Korn, B., Sedlacek, Z., Manca, A. *et al.* (1992) A strategy for the selection of transcribed sequences in the Xq28 region. *Hum. Mol. Genet.* **1**, 235–242.
- Bishop, J.O., Morton, J.G., Rosebach, M. & Richardson, M. (1974) Three abundance classes in HeLa-cell messenger-RNA. *Nature* **250**, 199–204.
- Galau, G.A., Klein, W.H., Britten, R.J. & Davidson, E.H. (1977) Significance of rare mRNA sequences in liver. *Arch. Biochem. Biophys.* **179**, 584–599.
- Milner, R.J., Lai, C., Lenoir, D. *et al.* (1986) Brain-specific gene expression. *Biochem. Soc. Symp.* **52**, 107–117.
- Ohlsson, R. (1989) Growth factors, protooncogenes and human placental development. *Cell. Differ. Dev.* **28**, 1–15.
- Patanjali, S.R., Parimoo, S. & Weissman, S.M. (1991) Construction of a uniform-abundance (normalized) cDNA library. *Proc. Natl Acad. Sci. USA* **88**, 1943–1947.
- Bird, A.P. (1986) CpG islands and the function of DNA methylation. *Nature* **321**, 209–213.
- Monaco, A.P., Neve, R.L., Coletti-Feener, C. *et al.* (1986) Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* **323**, 646–650.
- Liu, P., Legerski, R. & Siciliano, M.J. (1989) Isolation of human transcribed sequences from human rodent somatic cell hybrids. *Science* **246**, 813–815.
- Corbo, L., Maley, J.A., Nelson, D.L. & Caskey, C.T. (1990) Direct cloning of human transcripts with HnRNA from hybrid cell lines. *Science* **249**, 652–655.
- Kurnit, D.M. & Seed, B. (1990) Improved genetic selection for screening bacteriophage libraries by homologous recombination *in vivo*. *Proc. Natl Acad. Sci. USA* **87**, 3166–3169.
- Guo R.A., Neve, R.L., Klein, W.H., *et al.* (1993) Genomic scanning for expressed sequences in Xq21 identifies the glycerol kinase gene. *Nature Genet.* **4**, 367–372.
- Sargent, C.A., Affara, N.A., Bentley, E. *et al.* (1993) Cloning of the X-linked glycerol kinase deficiency gene and its identification by sequence comparison to the *Bacillus subtilis* homologue. *Hum. Mol. Genet.* **2**, 97–106.
- Adams, M.D., Kelley, J.M., Gocayne, J.D. *et al.* (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651–1656.
- Uberbacher, E.C. & Mural, R.J. (1991) Locating protein coding regions in human DNA sequences by a multiple-sensor-neural network approach. *Proc. Natl Acad. Sci. USA* **88**, 11261–11265.
- Freshney, R.I. (ed) (1987) *Culture of Animal Cells: A Manual of Basic Techniques* (2nd edn, Alan R. Liss, New York).
- Nisson, P.E., Rashtchian, A. & Watkins, P.C. (1991) PCR. *Meth. Appl.* **1**, 120.
- North, M., Sanseau, P., Buckler, A. *et al.* (1993) Efficiency and specificity of gene isolation by exon amplification. *Mammalian Genome* **4**, 466–474.
- Church, D., Stotler, C., Rutter, J. *et al.* (1994) Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* **6**, 98–105.
- Yaspo, M.-L., North, M. & Lehrach, H. (1993) Exon-enriched probe derived from a human chromosome 21 YAC by exon amplification. *Nucleic Acids Res.* **21**, 2271–2272.

- 37 Feinberg, A.P. & Vogelstein, B. (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6–13.
- 38 Litt, M. & White, R.L. (1985) A highly polymorphic locus in human DNA revealed by cosmid-derived probes. *Proc. Natl Acad. Sci. USA* **82**, 6206–6210.
- 39 Sealy, P.G., Whittaker, P.A. & Southern, E. (1985) Removal of repeated sequences from hybridization probes. *Nucleic Acids Res.* **13**, 1905–1922.
- 40 Dumas, J., Edwards, M., Delort, J. & Mallet, J. (1993) Oligodeoxyribonucleotide ligation to single-stranded cDNAs: a new tool for cloning 5' ends of mRNAs and for constructing cDNA libraries by *in vitro* amplification. *Nucleic Acids Res.* **19**, 5227–5232.

Chapter 18

Screening cDNA libraries
by transient expression in
mammalian cells

David L. Simmons

*Cell Adhesion Laboratory, ICRF, Institute of Molecular Medicine, John Radcliffe Hospital,
Headington, Oxford OX3 9DU, UK*

18.1	Introduction, 470	18.3.4	Screening for extracellular/secreted molecules by supernatant bioassay, 477
18.1.1	Basic outline of transient expression cloning, 470	18.4	Functional analysis of cDNA transfectants, 477
18.1.2	Advantages and disadvantages of transient expression cloning, 471		
18.2	cDNA library construction, 472		
18.2.1	cDNA synthesis and library construction, 472		
18.2.2	Vectors and the basis of transient expression methods, 472		
18.3	Screening methods, 476		
18.3.1	Introduction, 476		
18.3.2	Screening for surface molecules by panning and rescue, 476		
18.3.3	Screening for intracellular molecules by in situ labelling, 477		
			Protocol 96 cDNA library construction, 478
			Protocol 97 Screening for cell-surface proteins by transient expression, panning and episomal rescue, 486
			Protocol 98 Functional adhesion assays on cloned cDNAs transiently expressed in COS cells, 491
			References, 492

18.1 Introduction

Screening cDNA libraries by transient expression in mammalian cells has proved to be very effective for the isolation of cDNAs encoding secreted, surface and intracellular proteins. The first successful applications of transient expression cloning were in the field of growth factor research. In the mid-1980s, cDNAs encoding many cytokines, such as interleukin-3 (IL-3) [1] and interleukin-4 (IL-4) [2], were cloned by transient expression of cDNA libraries in COS cells, and screening of individual COS supernatants by a sensitive bioassay.

However, the single most successful application of transient expression screening was developed by Aruffo and Seed in 1987 [3–5]. It is based on transient expression of cDNA libraries in mammalian cells and rescue of specific cDNA clones by antibody capture and panning. The efficacy of this procedure has transformed the field of cell-surface clone isolation to such an extent that once a suitable antibody, ligand or cell line has been identified that recognizes a cell-surface molecule, the molecular cloning of the cDNA encoding it is now an essentially trivial process. Indeed, once one cell-surface molecule has been cloned, it is possible rapidly to clone the interacting ligand/receptor by using the extracellular domain of the first molecule, usually as an IgG1Fc chimaera, as an affinity reagent. This strategy has been used very successfully to clone leukocyte cell-surface molecules—for example, the CD40 ligand gp39 [6] and the Fas ligand [7].

Many orphan receptors (receptors with no known ligand) have been cloned by degenerate polymerase chain reaction (PCR)-based screens for tyrosine kinase or phosphatase domains. Once cloned, the extracellular domains of these orphan receptors can be used to identify and clone their cognate ligands. A good example of this strategy is the recent cloning

of the ligand for the haematopoietic Flt3/Flt2 receptor tyrosine kinase [8].

Since 1987, a large number of cell-surface molecules have been cloned using monoclonal antibodies to screen transiently expressed cDNA libraries (Table 18.1).

The technique has been extended to enable the cloning of intracellular proteins, though the number of successful examples of this strategy is still small [17,18].

Transient expression screens can also be used to clone genes by complementation of defective cell phenotypes. Expression of episomal-based cDNA libraries in these cells complements a defined defect allowing selection of the rescued cell. This type of screen has been particularly successful in the field of DNA repair defects. Many of the xeroderma pigmentosum (XP) mutations have been cloned by complementation of established XP cell lines. In addition the single genes defective in Fanconi’s anaemia [19] and paroxysmal nocturnal haemoglobinuria (PNH) [20] were also cloned by transient rescue.

In this chapter I shall describe the basics of cDNA library construction and then methods for transient expression screens for surface proteins, intracellular proteins and secreted proteins.

18.1.1 Basic outline of transient expression cloning

The essential elements of this technique are outlined in Fig. 18.1. It involves the construction of a representative cDNA library in a vector capable of replication and high-level expression in mammalian cells. After transfection of the library into the cell line, and transient expression of proteins encoded by it, the cells are screened in one of three different ways depending on the compartment where the protein of interest normally resides: intracellular,

Table 18.1 Some cell-surface proteins cloned by monoclonal antibody screening of expressed cDNA libraries.

T-cell adhesin/ activator CD2 [3] and its ligand LFA-3 (CD58) [5]
T-cell adhesin CD28 [4]
ICAM-1 (CD54) [9] and ICAM-3 (CD50) [10] recognizing LFA-1 (CD11a/CD18)
CD44 [11] recognizing hyaluronic acid
Endothelial intercellular adhesin CD31 [12]
Myeloid progenitor protein CD33 [13]
Haematopoietic progenitor sialomucin CD34 [14] which is a ligand for L-selectin
VCAM-1 (Cd106), an endothelial adhesin for VLA-4 on lymphocytes (this was cloned using a variation of the panning procedure employing cells directly as the recognition reagent [15])
ICAM-2 (CD102), an additional ligand for LFA-1, was cloned by using the ligand itself (LFA-1) as a direct panning reagent [16]

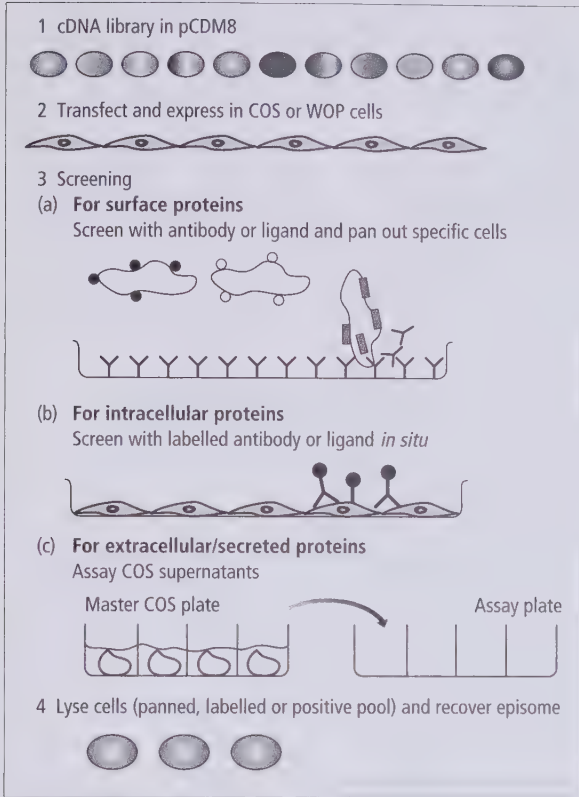


Fig. 18.1 Outline of transient expression screening.

surface or extracellular/secreted.

1 For intracellular proteins the cells expressing the library are fixed and dried *in situ* and screened with labelled ligand or antibody.

2 For surface proteins, a suspension of the cells is stained with the specific monoclonal antibody or ligand and then panned on plastic dishes coated with appropriate second antibody.

3 For secreted proteins, the cells expressing the library are divided into small pools and supernatants from these individual pools are assayed for bio-activity or antibody binding.

In all cases, the selected cells are lysed *in situ* and low molecular mass episomal DNA is recovered by differential precipitation (Hirt procedure; see Protocol 97). Episomes are then transformed into *Escherichia coli* and plated. This cycle of transfection–transient expression–selection–rescue usually needs to be repeated a further two or three times before individual recovered plasmids are analysed for expression of the specific protein.

18.1.2 Advantages and disadvantages of transient expression cloning

The main advantages of transient expression cloning systems are as follows.

1 **Rapidity** The transient expression profile reaches a maximum at 36–48 h after transfection. This means that each round of expression–selection and rescue only takes 3 days, so a complete 3- to 4-round library screen can be completed within 2–3 weeks.

2 **Full coding frame cDNA clones are isolated** By definition, only those cDNAs encoding the entire reading frame of the protein will be cloned. For surface proteins the cognate cDNA must at least have its ATG codon, extracellular domain, trans-membrane domain or lipid anchor, and stop-transfer sequence to give rise to a properly folded and processed surface molecule. In addition, as the selection is performed with monoclonal antibodies or direct ligands, the expressed molecule must be substantially the correct unmutated molecule.

3 **Functional studies on cloned surface molecules** The cloned cDNAs are in an efficient expression vector and can be used immediately for functional experiments such as radioligand binding quantification, cell adhesion studies, enzyme activity assay, etc.

The major disadvantages of transient expression cloning systems are the following.

1 **Multicomponent systems** A major limitation of transient expression cloning systems is encountered when dealing with multicomponent glycoprotein complexes where the expression of any individual component of the complex requires the expression of all other members of that system. Clearly, only single molecules can be cloned by this system and such complexes will be missed. This is a major defect, as many of the most important systems for cell recognition and signalling are multicomponent complexes. For example, the T-cell receptor (TCR) $\alpha\beta$ heterodimer requires expression of both chains to get either chain in the heterodimer to the surface. The TCR/CD3 complex δ -chain again requires multichain expression along with the TCR to get any surface expression of any of the CD3 chains. Integrins, major players in the process of cell–cell and cell–matrix adhesion, would also be missed by the expression cloning strategy as these are $\alpha\beta$ heterodimers where expression of the α -chain requires coexpression of the β -chain for surface presentation.

A way out of this cloning ‘black hole’ is the cotransfection of an existing expressing cDNA for one or all members of such complexes with the cDNA library under screen. For example, by cotransfection of integrin β -chains with cDNA libraries, it is possible to clone α -chains and vice versa.

It is also possible that the existing primate host cell integrins can act as ‘surrogate mothers’ for library-derived α - and β -chains. Primate α -chains could

associate with human β -chains giving rise to a species and chain heterodimer at the cell surface. Complementation of multichain complexes, by use of host cell-surface molecules may apply beyond the integrin family, but has not yet been tested.

2 Requirement for screening ligand A specific and high-affinity ligand is needed to screen the library. Usually, this has been a monoclonal antibody, although a variety of reagents—monoclonal antibodies, labelled ligands and whole cells—can be used to identify cell-surface molecules. Ligands in a labelled form—for example, iodinated interleukin-1 (IL-1) [21] and iodinated granulocyte-macrophage colony-stimulating factor (GM-CSF) [22]—have been used directly as affinity reagents to clone their cognate receptors.

Although the bias of this chapter is towards cell-surface molecules, the transient expression system has also been used extensively for the direct functional cloning of secreted molecules with biological activity such as cytokines and growth factors. IL-3, IL-4 and GM-CSF were cloned by this approach in the mid-1980s; supernatants from small pools of transfected COS cells were harvested and assayed in appropriate bioassays such as colony formation in soft agar, identification of positive pools, and repeat screening of positive pools. The cloning of all these cytokines led the way in transient expression technology. Cytokines are usually encoded by small highly abundant mRNAs so full-length cDNAs are likely to be well represented in cDNA libraries. In addition, the biological potency of many cytokines allows very small concentrations of active product to be detected in suitable mass assay systems.

Cell-surface molecules do not satisfy either of these two convenient criteria. Their genes are certainly not abundantly expressed; on average, most cell-surface molecules are present at 1000–50000 protein molecules per cell. The DNAs for most surface molecules are over 1 kb long and average around 1.5–2.5 kb. There is also the problem of a sensitive assay for detecting single clones in the library.

In the mid-1980s a high-efficiency transient expression vector system (pCDM8) was developed in the laboratory of Brian Seed at Harvard Medical School, Department of Genetics and Department of Molecular Biology, Massachusetts General Hospital, Boston. This allows the construction of representative cDNA libraries, high levels of expression and accumulation of cell-surface molecules ($1\text{--}5 \times 10^6$ molecules per cell surface). As CDM8 replicates to high copy number in COS cells, rescue and recovery of the episomal DNA is facilitated.

Coupled with the design of a variety of ingenious screening systems, this system has greatly increased our knowledge about cell-surface molecules by allowing the rapid cloning of cDNAs encoding them.

Clearly there are two parts to this technique: (i) construction of a cDNA library (see Protocol 96), and (ii) expression and selection of that library (see Protocols 97 and 98).

Protocols 96–98 derive in large part from the work of Brian Seed and Alexandro Aruffo.

18.2 cDNA library construction

18.2.1 cDNA synthesis and library construction

Methods for cDNA synthesis and library construction are given in Protocol 96. In this protocol I have described the basic routine procedures used to construct cDNA libraries, even though a number of ‘off the shelf’ rapid procedures are now available from a variety of vendors that allow essentially one-tube cells-to-poly(A)⁺ RNA preparation, usually involving oligo(dT) derivatized magnetic beads as the affinity isolation method. These are very quick and reliable methods but are extremely expensive, especially if several libraries are to be made over a period of time. The same applies to cDNA synthesis kits. Many are now available but if library construction is to be a routine part of laboratory skills, the cost of such an approach would be prohibitive.

The following list describes all the cDNA libraries constructed in my laboratory (all freely available from me, or now distributed by the UK Human Mapping Project Resource Centre, Sanger Centre, Hinxton Hall, Cambridge; see Appendix V for address), in the pCDM8 expression vector (Table 18.2). A large number of cell-surface molecules have been cloned from these libraries, and have been used by others for the isolation of many other genes by hybridization screens.

8.2.2 Vectors and the basis of transient expression methods

As with any library construction, the quality of the cDNA is of crucial importance to the isolation of any clones. The choice of vector into which the cDNA is ligated is linked to the choice of cell for expression. Molecular biologists have exploited elements of the genomes of mammalian DNA tumour viruses for vector construction and expression. The two essential elements of these viruses are:

- 1 origins of replication;
- 2 *trans*-acting DNA binding proteins that interact

Table 18.2 cDNA constructed in pCDM8.

<i>Human</i>
1 HPBALL (peripheral blood, acute lymphocytic leukaemia)
2 JY (lymphoblastoid, B EBV-positive)
3 HepG2 (hepatocellular carcinoma)
4 U937 (promonocytic leukaemia)
5 U937 (PMA-stimulated)
6 K562 (erythroleukaemia)
7 K562 (haemin-stimulated)
8 LAK (lymphokine-activated killer cells)
9 YT (HTLV-I-positive adult leukaemia, T cell)
10 HL60 (promyelocytic leukaemia)
11 HL60 (interferon-stimulated)
12 HT1080 (fibrosarcoma)
13 G361 amelanotic melanomas
14 C32 amelanotic melanomas
15 Placenta: full-term, normal pregnancy
16 Placental trophoblast (sorted 1st trimester)
17 Placental villi (1st trimester)
18 Human bone marrow (aspirate, ALL-positive, 1st remission)
19 HEL (human erythroleukaemia)
20 HUVEC (umbilical vein endothelial cell line)
21 HUVEC (stimulated with IL1- β (4 h))
22 HUVEC (stimulated with HT29 conditioned medium (48 h))
23 HUVEC (stimulated with DX3 conditioned medium (48 h))
24 L920 Hodgkin's lymphoma line
25 Fetal brain, 15–16 weeks
26 Normal colon
27 Colon carcinoma (solid tumour)
28 HT29 (colon carcinoma)
29 KG1 myeloblastic leukaemia
30 KG1A myeloblastic leukaemia
31 KG1B myeloblastic leukaemia
32 K562 (haemin-stimulated)
33 SU-DH-LI diffuse histiocytic lymphoma (non Hodgkin's lymphoma)
34 Mel DS1 amelanotic melanoma CD36-
35 Mel DS1 amelanotic melanoma (X-ray-induced)
36 Eosinophil
37 Fetal muscle
38 Natural killer cell
39 CEM (T cell)
40 Tonsil
41 HU-PC (phaeochromocytoma)
42 LAD (leukocyte adhesion deficiency type 1 patient EBV-B cells)
43 Normal human B cells (EBV-transformed)
44 DX3 melanoma
45 HCT116 colon carcinoma
<i>Rodent</i>
1 Mouse B cells (LPS)
2 Mouse T cells (ConA)
3 Mouse thymocytes
4 IC21 mouse macrophage cell line (PMA-stimulated)
5 Mouse spleen (NOD mouse)
6 Mouse bone marrow aspirate
7 Rat alveolar macrophage/ γ -IFN stimulated
8 Mouse serum stimulated macrophages

with the origin and the polymerase/primase complex to replicate the viral genome to high copy number in the appropriate cell line.

Two classes of virus have been exploited: papo-

vaviruses, especially simian virus 40 (SV40) and murine polyoma, and ebnaviruses, especially Epstein–Barr virus (EBV).

A crucial element of success in molecular cloning

by expression is the copy number of the virus-based vector in the host cell. First, it amplifies template per cell and thus increases the overall level of transcript production per cell; second, the amplified viral genome allows easy recovery from selected cells and thus reintroduction into *E. coli*.

This can best be illustrated by comparing vectors based on EBV with those based on SV40, as they differ markedly in their replication potential. EBV-based plasmids usually carry both the origin of replication (*oriP*) and the *trans*-acting origin amplifier (EBNA1), and can thus be expressed and amplified in any cell type. SV40-based plasmids usually only carry the origin of replication and have to be introduced into cell lines containing integrated copies of crippled SV40 genomes expressing the SV40 replicator protein large T antigen.

EBV-based plasmids only replicate to low copy number, typically 1–10 copies per cell, giving adequate levels of expression of specific molecules but posing difficult problems for subsequent recovery of those plasmids from selected cells. Indeed, higher levels of EBV episomes per cell are often toxic to the cell and are not tolerated. Typically, recovery of EBV episomes has had to make use of the very high efficiency of phage λ packaging extracts in order to rescue the vectors.

In contrast, SV40-based plasmids replicate to very high copy number per cell, typically 10^3 – 10^5 , yielding very high expression of specific molecules and also relatively facile recovery of the episomes from the selected cell. SV40 replicons are a burden to the cell in the long term and cells bearing them have elevated morbidity and eventual mortality. However, the burden can be supported for a sufficiently long time to allow expression selection and recovery of those cells.

For this reason, papovavirus-based plasmids have been the most widely used system for transient expression and rescue. Of the many types of papovaviruses, SV40 has been used most frequently [3,4], though the murine permissive virus, polyoma, has also been exploited [5].

There are many variants of SV40-based plasmids, and only a few will be described here. All share the same basic features: the SV40 origin of replication (a 350-bp fragment of the SV40 genome); a eukaryotic enhancer and promoter driving high level expression of the inserted cDNA or genomic fragment; downstream transcript processing elements (usually an intron and polyadenylation signal); and, finally, a prokaryotic origin of replication and some system for drug selection in *E. coli*.

p π H3M, developed by Aruffo and Seed in 1987, and pCDM8, developed by Seed in 1987, have been

successfully used for cloning many cell-surface molecules. pCDM8 postdates p π H3M, and has now superseded it. Consequently, pCDM8 will be described in detail.

Figure 18.2a illustrates the pCDM8 vector. It contains the powerful cytomegalovirus (CMV) enhancer and promoter driving expression of cDNA inserted at a polylinker cloning site flanked by nonpalindromic *Bst*XI sites. Downstream of this site is an intron and polyadenylation site, allowing efficient transcript processing and transport. pCDM8 contains both an SV40 origin of replication and a polyoma origin allowing replication of this vector in either primate cell lines such as COS-1 and COS-7 cells, and also murine polyoma-transformed lines such as WOP and COP. This is particularly useful if a specific monoclonal antibody cross-reacts with glycoproteins on the surface of COS cells, which are after all high primate cells of fibroblast/epithelial origin. Monoclonal antibodies raised in mice are highly unlikely to react with the surface of murine cells. The remaining elements of the vector allow replication in *E. coli* and drug selection mediated by a suppressor tRNA (*supF*) which suppresses amber stop codons in ampicillin- and tetracycline-resistance genes carried on a stable episome, p3, in the strain MC1061/p3. There is an M13 origin of replication, allowing production of single-stranded templates of the plasmid when appropriate F⁺ *E. coli* strains are superinfected with helper filamentous f1 phage such as M13. A T4 DNA promoter is included at the 5' edge of the cloning site to allow *in vitro* production of RNA templates for transcript terminus mapping and transcript production by T4 RNA polymerase.

Recently, some alternative versions of pCDM8 have been developed by commercial companies. The modifications have been of two types: (i) pcDNA1, which contains a slightly improved polylinker and addition of a 3' SP6 promoter for generation of antisense transcripts (Fig. 18.2b); (ii) pcDNA3, which is more radically altered by removal of the *supF* selection system and replacement with the β -lactamase ampicillin-resistance gene *Amp^r* (Fig. 18.2c). This allows selection of recovered plasmid in any highly competent *E. coli* strain capable of *ColE1* replication, and is not confined to the MC1061/p3 system. However, this latter vector has not yet been fully tested as an efficient platform for library construction and expression screening.

A different SV40-based expression vector pJFE14, has been constructed by John Elliott [23] (Fig. 18.2d). This uses the SR α promoter and the R and U 5' regions of the human T-cell lymphotropic virus I (HTLV I). An intron from the 16S RNA gene is placed

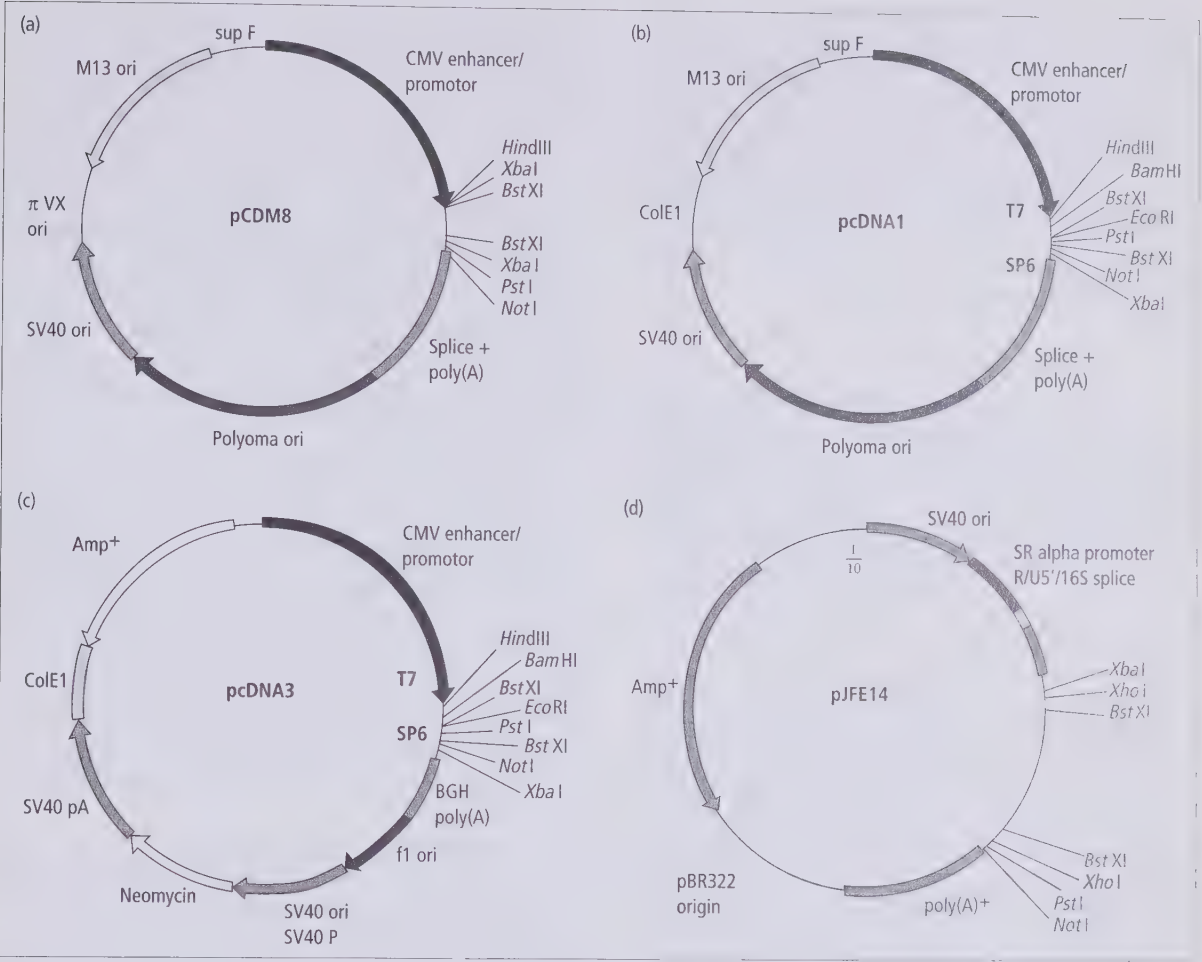


Fig. 18.2 Vectors. (a) pCDM8; (b) pcDNA1; (c) pcDNA3; (d) pJFE-14.

upstream of the *Bst*XI polylinker cloning site. The plasmid contains a ColE1 replicon and ampicillin resistance. Proponents of this vector argue that it overcomes some of the plasmid instability observed in the pCDM8/MC1061/p3 system.

All these vectors only replicate to high copy number in cells bearing SV40 or polyoma genomes. In 1981, Gluzman produced SV40-transformed African green monkey kidney fibroblasts (CV-1) cells bearing integrated copies of the SV40 genome, crippled by deletion of several bases at the SV40 origin. The resulting cells, COS-1 and COS-7, express high levels of SV40 large T antigen and permissivity factors, allowing high levels of SV40 replication per cell, but do not produce infectious viral genomes, so they are safe to work with under low-level containment. COS cells are eminently transfectable; with a DEAE-dextran/chloroquine regime (ref. 16, and see Section 18.3.2 below), it is routinely possible to achieve 50–60% of total

transfected cells expressing the introduced product. They have proved robust and reliable ‘work-horses’ for transient expression.

A useful future development would be the construction of SV40-based plasmids that also produced SV40 large T antigen, similar to the oriP/EBNA-1 p201-p205 system. Any cell line could then be transfected, irrespective of whether it contained endogenous SV40 genomes. This would allow genetic defects in defined cell lines to be complemented by introduced libraries, and rescue of the complementing episome.

Over the past 3–4 years EBV-based episomal cDNA libraries have been used very successfully to clone cDNAs by complementation. The advantage of the EBV system is that they carry their own trans-acting replication proteins (EBNA1) so can be expressed in any cell type. Many cell lines have been established from patients with inherited genetic defects such as XP, ataxia telangiectasia, Bloom’s

syndrome, Fanconi's anaemia, and PNH. Provided they can be robustly transfected, these cell lines can be used as recipients for wild-type cDNA libraries and restoration of normal phenotype can be screened and selected. Indeed, this technique has already been successfully applied to XP, Fanconi's anaemia [19] and PNH [20].

18.3 Screening methods

18.3.1 Introduction

I will now describe the methods for screening libraries by transient expression selection and rescue for proteins in each of three cellular compartments: intracellular, surface and extracellular/secreted. I will start with the method described in Section 18.3.2 and use it as the basis for the other two.

18.3.2 Screening for surface molecules by panning and rescue

This is described in detail in Protocol 97. cDNA libraries constructed as above in the pCDM8 vector are transfected into COS cells using DEAE-dextran as a facilitator [24] and chloroquine diphosphate to reduce lysosomal degradation of endocytosed DNA. Between 48 and 72 h after transfection, cells are lifted with phosphate buffered saline (PBS)/2 mM EDTA, washed in the same buffer containing 0.02% sodium azide and 5% fetal calf serum (FCS) at 4°C and incubated with monoclonal antibodies (mAbs) as tissue culture supernatants at minimal dilution ($\frac{1}{10}$ at most), at 4°C for 30 min, washed and applied to bacterial Petri dishes precoated with affinity-purified goat antimouse IgG. Cells are allowed to 'pan' for 1–2 h at room temperature and plates are then washed gently three to four times. It is possible to observe panned cells (10–100) per dish even at the first round of selection. However, on many occasions no cells may be seen if the clone being sought is of very low frequency in the library. Whether panned cells are observed or not at this stage, the procedure must be continued a further two rounds before a definitive assessment of success or failure is made. The panned cells are lysed *in situ* by applying Hirt squirt (0.8% SDS/10 mM EDTA). The cell lysate is harvested into an Eppendorf tube, 5 M NaCl is added and gently mixed and the tube is placed in a bath of wet ice for at least 1 h to allow precipitation of high molecular mass primate genomic DNA. The episomal DNA is recovered by spinning out the genomic DNA precipitate, followed by phenol extraction and ethanol precipitation. A fraction of the recovered

episomal DNA is transformed into highly competent MC1061/p3 and plated on LB agar containing ampicillin and tetracycline. A yield of 10^3 – 10^4 bacterial colonies should be obtained at this stage.

It is possible to continue to introduce the selected cDNA population into COS cells by DEAE-dextran facilitated transfection; however, a change of entry method is needed at this point. DEAE-dextran is a very efficient method of introducing DNA into cells, so it is an ideal method for the first round of screening to maximize representation of the cDNA library in COS cells. It is estimated that up to 10^3 – 10^4 different cDNA clones may be taken up by each COS cell by this method.

Consequently, a panned cell expressing the clone of interest will also contain 10^3 – 10^4 irrelevant cDNA clones, which will be represented in the yield of bacterial colonies from the first round. Thus, if DEAE-dextran was used for all subsequent rounds, a plateau of enrichment would be reached where the cDNA clone of interest would be contained within a heterogeneous population of irrelevant clones. This would mean that a very large number of individual bacterial colonies would have to be analysed by miniprep DNA isolation, individual transfection and mAb staining.

To prevent this, the second round of screening is initiated by introducing the bacteria into COS cells as spheroplasts or protoplasts (bacteria with cell walls removed). This is a very inefficient technique: only 1–5% of COS cells are transfected, a small number of protoplasts actually fuse with each COS cell, and each protoplast obviously contains only one cDNA clone. This means that a much smaller population of cDNA clones is introduced into each COS cell. The complexity of the resulting second round Hirt is thus greatly reduced and enrichment for the clone of interest is greatly enhanced.

The bacterial population is grown in liquid culture. Plasmids are amplified in the presence of spectinomycin, and converted to protoplasts by osmotic shock, EDTA chelation and lysozymal digestion. Protoplasts are introduced into COS cells by polyethylene glycol (PEG 1000 or PEG 1450) mediated membrane fusion. After another 36–48 h to allow transient expression of plasmid-encoded products, the COS cells are again incubated with the mAb, washed and panned.

As before, very few COS cells may be observed by visual scanning of the panning plates. Although considerable enrichment has occurred as a result of the first round of selection, the switch to a more inefficient method of transfection means that a similar number of COS cells will pan at this round.

A Hirt preparation is then made and processed in

the same way. Again, 10^3 – 10^4 bacterial colonies should derive from this round. A further round of protoplast fusion is needed before a definitive assessment of the success of the screening can be made. By this time, at the end of the third round of panning, COS cells should be visible on the panning dish.

Individual bacterial clones are picked, DNA isolated by standard SDS/alkaline lysis miniprep- aration methods (see refs 3, 4) and a fraction (10%) of that DNA is transfected into COS cells by DEAE-dextran facilitation. Forty-eight hours later the COS cells are stained *in situ* with the mAb, stained with a goat antimouse fluorescein isothiocyanate (FITC)-labelled second antibody, and scored by fluorescence microscopy. Results at this stage are rarely equivocal. Only a small number of individual colonies (10–20) need be analysed at this stage since 10–100% of these clones should be the clones of interest.

Much time and effort can be saved by screening with several mAbs at once as a pool. The pool of mAbs is used for the first two rounds of screening. At the third round the COS cells are incubated with each mAb separately and panned separately.

18.3.3 Screening for intracellular molecules by *in situ* labelling

This technique was developed by the group of Hans Clevers in Utrecht [17,18]. The cDNA library is transfected and expressed as for surface panning above. However, the COS cells are screened *in situ*, that is they are not lifted at days 2–3 post-transfection. In brief, the COS cells are rinsed in PBS and fixed in the culture dish for 10 min with methanol. All subsequent manipulations are performed at room temperature. The monolayer is washed twice with PBS and preincubated with PBS/5% FCS for 10 min, followed by a 1 h incubation with antibody or labelled ligand. Plates are washed twice with PBS followed by a 45-min incubation with peroxidase-labelled goat antimouse immunoglobulin, diluted 1:50 in 5% FCS/PBS. Peroxidase activity is subsequently visualized using a 5% dilution of a 4 mg ml⁻¹ stock of 9-amino-3-ethyl-carbazol in *N,N'*-dimethylformamide in 0.1 M NaAc

(pH 4.8) containing 0.1% H₂O₂ (leave in solution for 30–60 min). After washing with water, the plates are visually screened for positively stained (bright-red) cells with an inverted microscope. Positive cells are picked by scraping with a hand-held fine tip of a Gilson tip. Next, individual scraped cells are treated with Hirt squirt (see Protocol 97) and extracted. Plasmid DNA is transformed into MC1061/p3 and rounds of expression and selection are repeated as above.

18.3.4 Screening for extracellular/secreted molecules by supernatant bioassay

Again the basics of library transfection, transient expression and Hirt extraction are as already described for surface panning. The only difference comes in the actual screening procedure at 2–3 days post-transfection for each of the three rounds.

Twenty-four hours after transfection or protoplast fusion, the COS cells are trypsinized, pooled and counted in a haemocytometer. Cells are then aliquoted in appropriate pool sizes, usually 10^2 – 10^3 cells per well in either 24- or 96-well plates and allowed to adhere and express for a further 24–48 h. A fraction of the conditioned supernatant from each well is then harvested and applied to the assay plates. The assay will obviously be specifically designed for the protein being searched for. In the case of cytokines, growth factors or haematopoietic colony stimulating factors, a bioassay based on cell proliferation or colony growth or differentiation is the read-out. Positive wells are identified and a Hirt extract is made from the COS cells in the original master expression plate.

18.4 Functional analysis of cDNA transfectants

As described in Section 18.1.2, one of the advantages of the transient cloning system is that functional experiments on the cloned cDNA molecules can begin immediately. Transient expression of pure cDNA clones in COS cells can lead to the accumulation of up to 10^6 molecules per COS cell surface, so that functional assays can be performed directly on the cells (Protocol 98) [25].

Protocol 96 cDNA library construction

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Overview

- (a) RNA isolation
- (b) Poly(A)⁺ RNA preparation
- (c) cDNA synthesis
- (d) cDNA size fractionation
- (e) Ligation of cDNA to vector
- (f) General methods

(a) RNA isolation

This method is a simpler but more effective version of the original method of Chirgwin *et al.* [18]. It allows increased amounts of cell/tissue mass to be used and increased speed of preparation (much shorter centrifuge times).

Materials

- lysis buffer: guanidinium thiocyanate (GuSCN) (Fluka) in 25% lithium chloride (LiCl)
- 14.4 M β -mercaptoethanol
- RNase-free CsCl
- RNase-free water
- Falcon tubes
- Eppendorf tubes
- Beckman 5-ml polyallomer centrifuge tubes

Method

- 1 To each 1 ml of cell lysate dissolve 0.5 g guanidinium thiocyanate (GuSCN) (Fluka) in 0.58 ml of 25% lithium chloride (LiCl).
- 2 Filter through 0.45 μ m filter. Add 20 μ l of stock (14.4 M) β -mercaptoethanol.
- 3 In a 50-ml Falcon tube, centrifuge cells (1000–1500 *g*, 5 min). Or, for frozen tissue ground into a powder in dry-ice pellets, disperse pellet as a paste up the walls of the tube by banging.
- 4 Add 1 ml of the GuSCN/LiCl lysis buffer for up to 5×10^7 cells.
- 5 Shear the lysate immediately in a polytron homogenizer, top speed for 30–60 s, until DNA viscosity is completely gone. (Note: This is a very important step and cannot be overdone. It is vital to

completely shear the genomic DNA to avoid contamination of the RNA and also to avoid losses of RNA yield due to entrapment in the DNA layer at the GuSCN/CsCl interface on the gradient.)

- 6(a)** pipette 3.5 ml of the sheared lysate onto 1.5 ml of 5.7 M caesium chloride (CsCl) (RNase free; 1.36 g CsCl added to every 1 ml of 10 mM EDTA, pH 8.0) in a SW55 Beckman polyallomer centrifuge tube. Spin at 50 000 r.p.m. for 2 h.
- 6(b)** *For large-scale preparations (> 10⁸ cells)* Layer 25 ml lysate onto 12.5 ml of the 5.7 M CsCl cushion. Use a SW28 Beckman polyallomer centrifuge tube. Spin at 24 000 r.p.m. for 8 h.
- 7** At the end of the run, aspirate off the overlay through the CsCl interface and well down into the CsCl cushion, leaving only 1 ml in the bottom of the tube.
- 8** Aspirate off all residual liquid from the walls of the tube and scour a ring just above the remaining liquid level. Invert the tube and cut the tube just at the rounded part. Wipe off any liquid with a cottonbud or tissue.
- 9** Dissolve the clear RNA pellet in 0.4 ml of RNase-free water by triturating in a P1000 tip 10 times or more. Clear crystals of RNA should be visible that will eventually dissolve.
- 10** Pipette aqueous RNA into an Eppendorf tube.
- 11** Phenol extract (0.5 ml).
- 12** Chloroform extract (0.5 ml).
- 13** Add 10% vol. of 3 M sodium acetate and 2.5 vols ethanol. Place on dry ice for 10–15 min.
- 14** Spin in a minifuge (12 000 r.p.m.) for 5 min. Decant supernatant. Wash twice in 70% ethanol. Decant, remove residual ethanol with a P200 tip.
- 15** Redissolve RNA pellet in 0.5 ml of RNase-free water and titre (OD₂₆₀).
- 16** Store RNA at –70 °C.

(b) Poly(A)⁺ RNA preparation

Additional materials

- oligo(dT)-cellulose (Collaborative Research type IV)
- loading buffer (LB): 0.5 M lithium chloride, 50 mM Tris (pH 8.0), 5 mM EDTA, 1% SDS
- middle wash buffer (MWB): 100 mM LiCl, 50 mM Tris (pH 8.0), 5 mM EDTA, 1% SDS
- RNase-free sodium acetate (DEPC-treated) (Sigma)
- plastic disposable 10-ml column (Bio-Rad)
- Eppendorf tubes

Method

PREPARATION OF REAGENTS

- 1** Resuspend oligo(dT)-cellulose (Collaborative Research type IV) using 0.5 ml of dry powder per 1 ml of 0.1 M NaOH. Wash several times in RNase-free water.
- 2** Place in a plastic disposable 10-ml column, previously washed in 5 M NaOH and rinsed with water. Rinse oligo(dT) in 2–3 column vols of loading buffer (LB).

BINDING

- 3** Pour oligo(dT)-cellulose slurry into sterile 15-ml Falcon tube in 4–5 ml LB.
- 4** Heat total RNA, 1–2 mg at most, at 70 °C for 5 min. Chill on ice.
- 5** Adjust to 0.5 M with LiCl. Add to oligo(dT) slurry. Rotate on a wheel for 30 min.

WASHING

- 6** Decant into the disposable plastic column. Wash with 5 vols LB.
- 7** Wash with 5 vols of MWB.

ELUTION

- 8** Elute poly(A)⁺ RNA with serial 0.4-ml fractions of RNase-free water into Eppendorf tubes.
- 9** Add 10% by volume of RNase-free sodium acetate, 2.5 vols of ethanol and place on dry-ice for 30 min. Spin for 10 min.
- 10** Wash twice in room-temperature 70% ethanol.
- 11** Remove residual ethanol with a P200 tip and redissolve in 100 ml water.

Peak fractions from 1 to 2 mg starting total RNA should contain 20–50 µg pure, ribosomal RNA-free, poly(A)⁺ RNA. Fractions can be analysed on nonRNase-free ultrathin 1% agarose minigels (see Protocol 96f(i) below).

(c) cDNA synthesis

Additional materials

- mRNA prepared as in Protocol 96a
- RNase inhibitor (Boehringer)
- 1 M DTT

- linear polyacrylamide (LPA) carrier (see Protocol 96f(ii) below)
- RT1 buffer: 0.25 M Tris (pH 8.8), 0.25 M KCl, 30 mM MgCl₂
- RT2 buffer: 0.1 M Tris (pH 7.5), 25 mM MgCl₂, 0.5 M KCl, 50 mM DTT, 0.25 mg ml⁻¹ BSA molecular biology grade (Boehringer)
- oligo(dT) (dT₁₂₋₁₈) (Pharmacia)
- dNTPs (Pharmacia Ultrapure dGTP, dTTP, dATP, dCTP)
- reverse transcriptase (RT-XL, Life Sciences)
- DNA polymerase I (Boehringer)
- RNase H (Boehringer)
- low salt buffer (LSB): 60 mM Tris (pH 7.5), 60 mM MgCl₂, 50 mM NaCl, 2.5 mg ml⁻¹ BSA, 70 μM β-mercaptoethanol
- ligation additions (LA): 1 mM ATP, 20 mM DTT, 10 mM spermidine, 1 mg ml⁻¹ BSA, 100 mM MgCl₂
- 1× TE: 10 mM Tris, 1 mM EDTA pH 8.0
- *Bst*XI adaptors (kinased by T4 polynucleotide kinase or prepared with 5' phosphate on synthesis, see below) (Invitrogen)
- T4 DNA ligase (New England Biolabs)

Method

Double-stranded cDNA is constructed by a simplified 'one-tube' version of the original Gubler and Hoffman RNaseH method (see ref. 3).

FIRST STRAND

- 1** In a sterile Eppendorf tube add 5 μg mRNA. Heat to 100°C for 1 min. Quench on ice. Adjust volume to 70 μl with RNase-free water. Add: 20 μl 5×RT1 buffer; 2 μl RNase inhibitor (40 U μl⁻¹); 1 μl oligo(dT) (5 mg ml⁻¹) (dT₁₂₋₁₈); 2.5 μl dNTPs (25 mM) (dGTP, dTTP, dATP, dCTP); 1 μl DTT (1 M); 2 μl reverse transcriptase (the best, but unfortunately the most expensive, is Life Sciences RT-XL at 25 U μl⁻¹).
- 2** Incubate at 42°C for 40 min. Heat inactivate at 70°C for 10 min.

SECOND STRAND

- 3** To the same tube, add: 320 μl RNase-free water, 80 μl RT2 buffer, 5 μl DNA polymerase I (5 U μl⁻¹), 2 μl RNase H (2 U μl⁻¹).
- 4** Incubate at 15°C for 1 h.
- 5** Switch tube to room temperature for a further hour.
- 6** Stop reaction by adding 20 μl 0.5 M EDTA, pH 8.0.
- 7** Phenol extract (add 0.5 ml phenol, vortex, spin, remove aqueous phase).
- 8** Chloroform extract (0.5 ml).
- 9** Precipitate by adding 10% vol. of 5 M NaCl, LPA carrier to 20 μg ml⁻¹ and adding 2 vols ethanol.

10 Place on dry ice pellets for 10 min. Spin for 2–3 min only. Wash twice with room-temperature 70% ethanol. Remove residual ethanol with a P200 tip.

11 Redissolve cDNA pellet in 240 µl water.

I have found that addition of T4 DNA polymerase at the end of second-strand synthesis, to blunt the cDNA ends, does not appreciably or reliably increase the yield of ligatable cDNA so it is simply omitted.

LIGATION OF ADAPTORS

12 To the 240 µl of cDNA, add: 30 µl 10× LSB, 30 µl 10× LA, 5 µg equimolar mixture of the *Bst*XI adaptors (kinased by T4 polynucleotide kinase or prepared with 5' phosphate on synthesis, see Protocol 96f below), 1 µl T4 DNA ligase (400 U µl⁻¹).

13 Incubate overnight at 15 °C.

14 Phenol extract, chloroform extract and ethanol precipitate as above.

15 Resuspend final cDNA pellet in 100–200 µl TE.

*Bst*XI adaptors are available commercially. Directional cloning of cDNA is possible using an oligo(dT) primer containing a *Not*I site for first strand synthesis, ligating *Eco*RI adaptors to the second strand, cutting with *Not*I and ligating the cDNA into *Eco*RI-*Not*I vector. While directional cloned cDNA is obviously an advantage for expression cloning, this system has proved to be very inefficient (probably due to the inefficiency of the *Not*I) and overall yields are then much below what could be achieved by the non-directional *Bst*XI adaptors, so negating the advantage of 100% correct orientation with respect to the vector enhancer/promoter.

(d) cDNA size fractionation

The best way we have found for achieving the dual goals of efficient non-ligated adaptor removal and size fractionation of the cDNA is kinetic density centrifugation on continuous gradients of 5–20% potassium acetate.

Additional materials

- potassium acetate (KOAc)
- 5-ml gradient maker (Hoeffer SM5)

Method

1 Prepare continuous linear gradients in a 5-ml gradient maker. Add 2.5 ml 20% KOAc to the back chamber. Add 2.5 ml 5% KOAc to the

front chamber. Fill a Beckman 5-ml SW55 polyallomer centrifuge tube with the 5 ml 5–20% continuous KOAc gradient.

- 2 Layer the 100–200 ml cDNA very gently onto the top of the gradient. Spin at 50 000 r.p.m. for 3–4 h. Puncture the tube near the bottom with a 21-gauge butterfly needle and collect 0.4-ml fractions.
- 3 Add 5 µg of LPA carrier, 2 vols of ethanol and freeze on dry ice for 10 min. Spin for 3 min, wash twice with room-temperature 70% ethanol. Remove residual ethanol with a P200 tip. Resuspend each fraction in 20 µl water.
- 4 Analyse 2 µl of each fraction on a 1% agarose minigel (see Protocol 96f(i)). Pool fractions with cDNA larger than 500–750 bp. Fractions can be kept separate for each size band — for example, 500–1000 bp, 1000–1500 bp, 1500–2000 bp, 2–3 kb, 4 kb and larger — resulting in five pools of a tight size range and ligated separately to vector to make very discrete size-range libraries.

(e) Ligation of cDNA to vector

Additional materials

- cDNA as prepared in Protocol 96d
- vector pCDM8
- *E. coli* MC1061/p3
- LB agar plates containing 10 µg ml⁻¹ ampicillin and 10 µg ml⁻¹ tetracycline
- 10-cm bacterial Petri dishes
- 24 × 24 cm culture dishes

Method

SMALL-SCALE TEST LIGATIONS

- 1 Use 1–5% vols of the cDNA. Ligate to a constant amount (10–20 ng) of vector (pCDM8, cut with *Bst*XI and the stuffer fragment removed by KOAc gradient centrifugation as above for the cDNA). Ligations are in a small volume (10–20 µl) with 10 ng of vector for 1 h at room temperature.
- 2 Transform 10% of the ligation mix (1–2 µl) into 50 µl ‘super-competent’ MC1061/p3 cells (see Protocol 96f(iv) below).
- 3 Place on ice for 15 min, heat shock at 37 °C for 5 min.
- 4 Plate on 10-cm LB agar plates containing ampicillin at 10 µg ml⁻¹ and tetracycline at 10 µg ml⁻¹, with a 5 ml LB agar overlay poured during the heat-shock incubation to provide a ‘drug-free zone’ for the cells to grow and express drug-resistance genes before being exposed to the antibiotics. By using 1% of the cDNA and 10% of the ligation mix, the number of colonies per plate on this small scale is multiplied by 10³.

The key quality control checks on any library are primary complexity and insert size range.

- A library size of 2×10^5 – 10^6 colonies should be aimed for; anything less is unsatisfactory.
- Insert size range should be 1–2 kb, with 95% of colonies containing inserts. Standard alkaline/SDS lysis miniprep DNA (see, for example, ref. 18) should be used to analyse the inserts in at least 20–30 colonies.

LARGE-SCALE LIGATIONS

If both of the above criteria are satisfied (library size and insert size range), proceed to large-scale ligation using most if not all of the cDNA and proportionately more vector. The entire cDNA yield should consume no more than 1–2 g of the *Bst*XI-cut (stuffer minus) purified vector.

- 1 Ligate as above and transform into competent cells, ensuring that the ligation mix is kept at less than 2–4% of competent cell volume (spermidine in the ligation buffer severely inhibits transformation efficiency).
- 2 Plate on 24 × 24 cm dishes at 10^5 colonies per plate. Harvest the resulting primary plating and maxiprep using alkaline/SDS lysis (see ref. 18) and caesium chloride density gradients.
- 3 Store cDNA library as DNA at -20°C .

I have experienced no deterioration of library stocks over the 6 years I have been making them. Also, cDNA libraries can be safely amplified by re-transformation of the primary library stocks without gross loss of library complexity.

(f) General methods

Materials

- 1% agarose
- TAE running buffer
- blood agglutination slides
- 5% acrylamide solution with ammonium persulphate (0.1%) and TEMED (0.1%)
- kinasing buffer (KB): 0.5 M Tris (pH 7.5), 10 mM ATP, 20 mM DTT, 10 mM spermidine, 1 mg ml⁻¹ BSA, 100 mM MgCl₂
- TYM agar: 2% Bacto-Tryptone, 0.5% yeast extract, 0.1 M NaCl, 10 mM MgSO₄
- transformation buffer I (TFBI): 30 mM potassium acetate, 50 mM MnCl₂, 100 mM KCl, 10 mM CaCl₂, 15% glycerol (v/v)
- transformation buffer II (TFBII): 10 mM Na-MOPS (pH 7.0), 75 mM CaCl₂, 10 mM KCl, 15% glycerol

(i) *Ultrathin 1% agarose minigels* These are prepared on 'old-style'

blood agglutination slides (Blue Star microslides 76×51 mm). Twenty to thirty slides are made at once, laid out on Parafilm with fine teeth combs (6–7 wells per slide). Pipette 8–10 ml of warm (50°C) molten 1% agarose in TAE running buffer. The agarose is held by surface tension as a bubble.

These gels hold 10 µl per well and can be run extremely fast (200V, 15 min), allowing rapid easy monitoring of all the steps of cDNA synthesis procedure. They are extremely thin and have very low autofluorescence background allowing 10–50 ng cDNA to be readily visualized by trans-UV illumination.

- (ii) *Linear polyacrylamide carrier* Linear polyacrylamide (LPA) has proven to be a reliable and completely noninjurious inert carrier allowing efficient precipitation of picogram quantities of DNA at near zero cost. There is no risk of contamination with tRNA or rRNA and conversion to cDNA during cDNA synthesis reactions.

Prepare by polymerization of a 5% acrylamide solution with ammonium persulphate (0.1%) and TEMED (0.1%). No *bis*-acrylamide is present, so only linear chains of polyacrylamide form. This solution is 50 mg ml⁻¹ and a working solution at 2 mg ml⁻¹ is diluted from this. This is stored at -20°C and may be frozen and thawed many times. Usually, 5–10 µg per precipitation reaction is sufficient.

- (iii) *Adaptor preparation* Adaptors are added enzymatically using polynucleotide kinase:

- adaptors at 1 mg ml⁻¹ in 50 ml reaction volume;
- 5 ml of 10×KB;
- 20 units of T4 polynucleotide kinase.

Incubate at 37°C overnight. The non-self-compatible *Bst*XI adaptors are 5'-CTTTAGAGCACA-3' and 5'-CTCTAAAG-3'.

NB It is essential that the adaptors are efficiently purified by high pressure liquid chromatography (HPLC) before use and that each new batch is tested on an existing batch of 'good' cDNA. Good adaptors are one of the keys to good library construction.

- (iv) *Super-competent cells* Many protocols exist for making bacterial cells competent for transformation. We have used a simple two-step chemical method, which allows the production of cells with a competency of 1–5×10⁸. This level satisfies the dual need for cDNA library transformations and amplification of recovered episomal DNA from library screens (see below).

- Streak out *E. coli* MC1061/p3 on a fresh TYM plate. Incubate overnight. Pick single colonies into 5 ml of TYM, grow on a wheel with good aeration for 3–4 h.
- Dilute to 100 ml in a 250-ml flask in TYM, grow to mid-log OD₆₀₀=0.5. Dilute to 500 ml in a 2-litre flask in TYM, grow to mid-log OD₆₀₀=0.5. Rapidly chill cultures by swirling in water/ice.
- Pellet bacteria in a Beckman J6 centrifuge in 1-litre pots at 4000 r.p.m. for 15 min. Resuspend pellet very gently and slowly in 100 ml TFB1 on ice/water. Pellet at 2500 r.p.m. for 10 min at 4°C.

- Resuspend pellet in 20 ml TFBII. Aliquot in prechilled Eppendorf tubes and flash freeze in liquid nitrogen. Store at -70°C . Competency can be tested on supercoiled plasmid standard stocks at 100, 10, 1, and 0.1 pg levels or by relative comparisons to existing tested batches of ligated cDNA or library screen Hirts (Protocol 97). Competent cells maintain the desired level of competency for at least 3–6 months.

.....

Protocol 97 Screening for cell-surface proteins by transient expression, panning and episomal rescue

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Overview

- (a) Round 1: DEAE-dextran transfection, expression, screening, panning, episomal rescue and transformation
 - (b) Round 2: protoplast fusion, expression, screening, panning, episomal rescue and transformation
 - (c) Round 3: repeat of round 2 plus scoring individual clones
- (a) Round 1: DEAE-dextran transfection, expression, screening, panning, episomal rescue and transformation

Materials

- COS cells
- DME/10% FCS
- cDNA library
- *E. coli* strain MC1061/p3
- 1×TE
- DEAE-dextran (Sigma, M_r 400 000)
- chloroquine diphosphate
- NuSerum (Collaborative Research) or Ultrosor G (Gibco-BRL)
- osmotic shock medium: PBS/10% dimethyl sulphoxide (DMSO)
- panning buffer (PB): PBS, 2 mM EDTA, 0.1% sodium azide, 5% FCS
- LB medium containing 10 µg ml⁻¹ ampicillin and 10 µg ml⁻¹ tetracycline
- monoclonal antibodies (mAbs)
- affinity-purified goat antimouse IgG
- spectinomycin
- Hirt squirt (0.8% SDS/10 mM EDTA)
- Falcon 15-cm culture dishes
- 10-cm Petri dishes
- preprepared panning plates (see stage 6)

Method

TRANSFECTION

- 1** Grow COS cells in DME/10% FCS at 50–75% confluency (most conveniently in Falcon 15-cm Intergrid culture dishes). Use 10–20 μg cDNA library in pCDM8, or a similar SV40-based vector, to transfect 1×10^7 COS cells using 400 $\mu\text{g ml}^{-1}$ DEAE-dextran as a facilitator [24], and chloroquine diphosphate at 100 μM . Dilute cDNA library DNA well below 1 mg ml^{-1} in TE, add DEAE-dextran and dilute up in medium either without serum or with 10% NuSerum or a low protein concentration serum supplement such as Ultrosor G. Alternatively, cells can be transfected in medium alone, although some increased mortality will occur. Leave on for up to 4 h, or until the COS cells begin to look vacuolated.
- 2** Aspirate medium, and add 15 ml PBS/10% DMSO osmotic shock medium for 2 min. Aspirate, and replace with regular medium.
- 3** Twenty-four hours after transfection, trypsinize cells and replate on fresh culture dishes to remove residual adsorbed DEAE-dextran. It is essential to do this in order to be able to lift the COS cells with EDTA the following day and achieve a monodisperse single cell suspension.

SCREENING

- 4** Forty-eight to 72 h after transfection, aspirate medium, wash twice with PBS only and lift cells with 10 ml PBS containing 2 mM EDTA. Put dishes at 37°C for 10–15 min.
- 5** Wash in PB at 4°C. Incubate with mAbs (either neat or, at most, a 1 : 10 dilution of tissue culture supernatants or a 1 : 100–1 : 1000 dilution of ascites or 1 $\mu\text{g ml}^{-1}$ purified antibody), at 4°C for 30 min. Wash twice in cold PB.

PANNING

- 6** Preparation of panning plates: Coat 10-cm Falcon Petri dishes with 5 ml of a 10 $\mu\text{g ml}^{-1}$ solution of affinity-isolated goat antimouse IgG in 50 mM Tris (pH 9.5) for 1–2 h. Wash three times in PBS. Block remaining sites by overnight incubation with 5 ml per dish of blocking buffer (PBS, 2 mg ml^{-1} BSA). Aspirate blocking buffer, and store plates at –20°C for up to 6 months.
- 7** Apply the antibody-labelled cells to the prepared panning plates. Leave in a vibration-free part of the laboratory for 2–3 h to allow cells to pan gently at room temperature. Wash panning plate very gently using a pipette only, not a suction line. Remove cells, and add 5 ml PB to one edge of the dish held at a 30° angle. Gently roll around two to three times and remove the PB from the opposite edge of the dish. Repeat washing three to five times.

- 8** Check the efficiency of washing under an inverted microscope. Gently roll the dish on the microscope stage and check for the general number of free-floating cells still remaining. Continue washing until no floaters remain. Assess the level of panned cells and whether there are large numbers of obviously dead cells non-specifically stuck to the dish.

EPISOMAL RESCUE

- 9** *Preparation of Hirt* Lyse the specifically panned cells *in situ* in 400 μ l Hirt squirt (0.8% SDS/10 mM EDTA). Gently swirl around the dish to cover efficiently.
- 10** Cut 1–2 mm from the end of a Gilson P1000 tip and pipette the lysate gently into an Eppendorf tube (this avoids shearing the genomic DNA). Add 100 μ l 5 M NaCl, mix gently by inversion and place in a bath of wet ice for at least 1 h to allow precipitation of high molecular mass primate genomic DNA.
- 11** Recover episomal DNA by spinning out the white genomic DNA precipitate in minifuge for 5 min.
- 12** Remove clear supernatant to a fresh tube and respin if any part of the precipitate carries over. Remove clear supernatant to a fresh tube.
- 13** Add 0.5 ml phenol, vortex for 1 min, spin and remove aqueous phase to a fresh tube.
- 14** Extract with 0.5 ml chloroform, spin and remove aqueous phase to a new tube.
- 15** Add 5 μ g of LPA carrier (see Protocol 96f(ii) for recipe), mix. Add 2 vols ethanol, mix and place on dry ice for 10 min. Spin for 3 min, wash twice with 70% ethanol, remove residual ethanol with a P200 tip and dissolve Hirt in 50 μ l TE.

TRANSFORMATION OF HIRT

- 16** Ten to 30% (5–15 μ l) of the recovered episomal DNA is transformed into 0.5 ml highly competent MC1061/p3 (highly competent is more than 10^8 colonies per μ g). Plate on one 24 \times 24 cm LB agar plate containing ampicillin and tetracycline each at 10 μ g ml⁻¹. A yield of 10^3 – 10^4 bacterial colonies should be obtained for the first round Hirt. Anything less is a failure, so start again. Anything more is a bonus, so continue.

(b) Round 2: Protoplast fusion, expression, screening, panning, episomal rescue and transformation

Additional materials

- trypsin
- lysozyme
- DME/10% sucrose/10 mM MgSO₄
- PEG 1000 or PEG 1450/50% DME
- gentamicin sulphate
- laminar flow hood

Method

EXPANSION AND AMPLIFICATION OF ROUND 1 POPULATION

- 1** Scrape the bacterial population from round 1 into a slurry in 20–50 ml LB medium containing ampicillin and tetracycline, both at 10 µg ml⁻¹. Titre a 1 : 10 or 1 : 100 dilution at OD₆₀₀.
- 2** Grow this population in 100–200 ml liquid culture with vigorous shaking at 37 °C from a starting inoculum of OD₆₀₀ = 0.1 to OD₆₀₀ = 0.5. Amplify plasmids by overnight incubation, with shaking, in the presence of 100 µg ml⁻¹ spectinomycin. This allows some amplification of plasmid copy number per bacterium, and also arrests bacterial growth so that the fusion inoculum is not excessive.
- 3** Prepare COS cells now for protoplast fusions the next day. Trypsinize COS cells (see Chapter 8, Protocol 32 or ref. 26) and plate at 50–75% confluency in 10-cm culture dishes; you will need two 10-cm dishes per 100 ml of bacterial culture.

CONVERSION TO PROTOPLASTS

The overnight bacterial liquid culture is converted to protoplasts by sequential osmotic shock, EDTA chelation and digestion with lysozyme.

- 4** Pellet bacteria by centrifugation (e.g. Beckman JA14/GSA rotor, 250 ml bottles) for 5 min at 10 000 r.p.m. Resuspend the bacterial pellet in 5 ml cold 20% sucrose, 50 mM Tris (pH 8.0). Add 1 ml of lysozyme (10 µg ml⁻¹) freshly dissolved in 250 mM Tris (pH 8.0).
- 5** Incubate at 4 °C for 5 min.
- 6** Add 2 ml cold EDTA (0.25 M), pH 8.0.
- 7** Incubate at 4 °C for 5 min.
- 8** Add 2 ml Tris (50 mM), pH 8.0.
- 9** Incubate at 4 °C for 5 min.
- 10** Place in a 37 °C waterbath for 5 min. Place on ice and check for percentage conversion to spheroplasts by microscopy. (There should be 90% conversion of rod-shaped bacteria to spherical protoplasts.)

PROTOPLAST FUSION

Perform all manipulations in a laminar flow hood.

- 11** Add 2 ml cold DME/10% sucrose/10 mM MgSO_4 slowly, dropwise, from a 25-ml pipette, swirling all the time. Remove media from 10-cm dishes of COS cells at 50–75% confluency. Add 15 ml of the spheroplast slurry to each dish. Place dishes in bottom of buckets of bench-top centrifuge (Beckman GPR, Sorval RC6000) with rubber bases still in. Two dishes per bucket can be accommodated.
- 12** Spin at 2500 r.p.m. for 10 min at 4°C and decelerate without brake to avoid disruption of protoplast skins.
- 13** Aspirate fluid from dishes. Add 5 ml of 50% (w/w) PEG 1000 or PEG 1450/50% DME into the centre of the dish. After the PEG has been added to last dish, prop all the dishes up on their lids so that the PEG drains to the bottom edge.
- 14** Aspirate PEG layer. Leave for fusion to occur over 90–120 s (PEG 1000) or 120–150 s (PEG 1450). Stop fusion by adding 5 ml DME into the centre of the dish. The PEG layer will be swept radially away by the medium.
- 15** Aspirate and repeat the washing. Aspirate and add 10 ml of DME/10% FCS containing $10 \mu\text{g ml}^{-1}$ gentamicin sulphate, and leave for 4 h, over which time the protoplast layer will gradually disintegrate.
- 16** Swirl the dishes to disrupt the protoplast skins, aspirate and change the medium. Gentamicin sulphate is essential for these cultures because the residual bacterial population is so massive that penicillin and streptomycin are completely ineffective.

EXPRESSION, SCREENING AND PANNING

Leave fused COS cells for 36–48 h to allow transient expression of plasmid-encoded products. Repeat mAb screening and panning as above for round 1. Prepare Hirt DNA (see a, stage 9 above), extract, precipitate and transform in MC1061/p3. The yield should be 10^3 – 10^4 bacterial colonies.

- (c) Round 3: repeat of round 2 plus scoring individual clones

Additional materials

- FITC-labelled goat antimouse antibody (Sigma)

Method

- 1** Perform a further round of protoplast fusion as above. At the end of this round, transform 10% of the final Hirt DNA into 50 μl competent MC1061/p3 cells, and plate on a 10-cm Petri dish of LB + ampicillin and tetracycline as above. Incubate overnight.

- 2 Pick 10–30 individual bacterial colonies each into 2.5 ml of LB + amp/tet and grow at 37 °C with vigorous shaking to saturation (8h minimum culture time). Isolate plasmid DNA by standard alkaline/SDS lysis methods (see ref. 26).
- 3 Transfect 10–30% of the miniprep DNA into COS cells in 6-well cluster plates by DEAE-dextran facilitation protocol (see (a), stage 1 above). Forty-eight hours later, screen the COS cells *in situ* with the mAb at 4 °C for 30 min, wash three times and stain with a 1 : 100 dilution of FITC-labelled goat antimouse second antibody. Wash three times more and fix with PBS/2% formaldehyde.
- 4 Score individual wells by fluorescence microscopy. The percentage of positive clones can vary from 10 to 100% depending on many variables, including how abundant the original cDNA was in library, the affinity of the antibody or ligand, and the overall efficiency of the three rounds of expression, panning and rescue.

.....

Protocol 98

Functional adhesion assays on cloned cDNAs transiently expressed in COS cells

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

Materials

- COS cells
- trypsin
- cloned cDNAs prepared as in Protocols 96 and 97
- DEAE-dextran
- test cells for adhesion as appropriate

Method

- 1 Trypsinize COS cell stocks and re-plate at a density of $1 \times 10^4 \text{ cm}^{-2}$. Transfect 10–20 µg plasmid DNA into the COS cells 4 h at 37 °C using the DEAE-dextran method (see Protocol 97a, stage 1). Leave cells for a further 18h, trypsinize and replat at a density of 10^4 cm^{-2} on your chosen assay format: 6-well, 24-well, or 96-well plates, or 3-cm or 6-cm dishes. It is essential to ensure that the cell density is correct and that the distribution is even throughout the well or dish. To avoid cells ‘piling-up’ in the centre of the plate, gently rock the dishes every 1–2h for 6h after replating to redistribute the cells.

Transient expression of the encoded cDNA can be measured 48 h after transfection, but is optimal if left for 72 h. (This is especially true for double transfections, e.g. expression of two subunits of a dimeric

receptor.) Cell-surface molecule expression can be monitored by immunocytochemistry using specific monoclonal antibodies, or by functional adhesion assay. Functional adhesion assays can be performed using radioisotopically labelled cells (overnight incorporation of [³H]thymidine) or unlabelled cells coupled with visual assessment of adhesion photomicroscopically after fixing and staining COS cell/test cell rosettes in 0.2% crystal violet in 10% phosphate-buffered formalin (pH 7.4).

- 2 Allow test cells to adhere to COS transients for up to 1 h. Wash three to five times, monitoring for removal of floating cells. Fix in PBS/2% formaldehyde. Fixed cell rosettes can be directly visualized and photographed under phase contrast using an inverted microscope.

References

1 Lee, F., Yokota, T., Otsuka, T. *et al.* (1986) Isolation and characterization of a mouse interleukin cDNA clone that expresses B-cell stimulatory factor 1 activities and T-cell and mast cell stimulating activities. *Proc. Natl Acad. Sci. USA* **83**, 2061–2065.

2 Yang, Y., Ciarletta, A.B., Temple, P.A. *et al.* (1986) Human IL-3 (Multi-CSF): identification by expression cloning of a novel hematopoietic growth factor related to murine IL-3. *Cell* **47**, 3–19.

3 Seed, B. & Aruffo, S. (1987) Molecular cloning of the human CD2 antigen by a rapid immunoselection procedure *Proc. Natl Acad. Sci. USA* **84**, 3365–3369.

4 Aruffo, S. & Seed, B. (1987) Molecular cloning of a CD28 cDNA by a high efficiency COS cell expression system. *Proc. Natl Acad. Sci. USA* **84**, 8573–8577.

5 Seed, B. (1987) An LFA-3 cDNA encodes a phospholipid-linked membrane protein homologous to its ligand CD2. *Nature* **329**, 840–842.

6 Armitage, R.J., Fanslow, W.C., Strokbine, L. *et al.* (1992) Molecular and biological characterization of a murine ligand for CD40. *Nature* **357**, 80–83.

7 Suda, T., Takahashi, T., Golstein, P. & Nagata, S. (1993) Molecular cloning and expression of the Fas ligand, a novel member of the tumour necrosis factor family. *Cell* **75**, 1169–1178.

8 Lyman, S.D., James, L., Vanden Bos, T. *et al.* (1993) Molecular cloning of a ligand for the flt3/flk-2 tyrosine kinase receptor: a proliferative factor for primitive hematopoietic cells. *Cell* **75**, 1157–1167.

9 Horst, E., Wijngaard, P.L.J., Metzelaar, M. *et al.* (1991) A method for cDNA cloning in COS cells irrespective of the subcellular site of expression. *Nucleic Acids Res.* **19**, 4556–4559.

10 Metzelaar, M. Wijngaard, P.L.J., Peters, P.J. *et al.* (1991) CD63 antigen, a novel lysosomal membrane glycoprotein, cloned by a screening procedure for intracellular antigens in eukaryotic cells. *J. Biol. Chem.* **266**, 3239–3245.

11 Strathdee, C.A., Gavish, H., Shannon, W.R. & Buchwald, M. (1992) Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature* **356**, 763–767.

12 Miyata, T., Takeda, J., Iida, Y. *et al.* (1993) The cloning of PIG-A, a component in the early step of GPI-anchor biosynthesis. *Science* **259**, 1318–1320.

13 Sims, J.E., March, C.J., Cosman, D. *et al.* (1988) cDNA expression cloning of the IL-1 receptor, a member of the immunoglobulin superfamily. *Science* **241**, 567–569.

14 Gearing, D., King, J., Gough, N. & Nicola, N. (1989) Expression cloning of a receptor for human granulocyte-macrophage colony-stimulating factor. *EMBO J.* **8**, 3667–3676.

15 Elliott, J.F., Albrecht, G.R., Gilladoga, A. *et al.* (1990) Genes for *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **87**, 6363–6367.

16 Gorman, C. (1986) High efficiency gene transfer into mammalian cells. In *DNA Cloning* (Glover, D.M., ed.), pp. 143–190 (IRL Press, Oxford).

17 Simmons, D. & Needham, L.A. (1992) Cloning cell-surface molecules using monoclonal antibodies. In *Vascular Endothelium: Interactions with Circulating Cells* (Gordon, J., ed.), pp. 3–29 (Elsevier, Amsterdam).

18 Sambrook, J., Fritsch, E.F. & Maniatis, T. (eds). (1989) *Molecular Cloning. A Laboratory Manual* (2nd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

19 Simmons, D., Makgoba, W. & Seed, B. (1988) ICAM-1, an adhesion ligand of LFA-1, is homologous to the neural cell adhesion molecule NCAM. *Nature* **331**, 624–627.

20 Fawcett, J., Holness, C.L., Needham, L. *et al.* (1992) Molecular cloning of ICAM-3, a third ligand for LFA-1, constitutively expressed on human leukocytes. *Nature* **360**, 481–484.

21 Stamenkovic, I., Amiot, M., Pesando, J. & Seed, B. (1989) A lymphocyte molecule implicated in lymph node homing is a member of the cartilage link protein family. *Cell* **56**, 1057–1062.

- 22 Simmons, D., Walker, C., Power, C. & Pigot, R. (1990) Molecular cloning of CD31, a putative intercellular adhesion molecule closely related to carcinoembryonic antigen. *J. Exp. Med.* **171**, 2147–2152.
- 23 Simmons, D. & Seed, B. (1988) Isolation of a cDNA encoding CD33, a differentiation antigen of myeloid progenitor cells. *J. Immunol.* **141**, 2791–2800.
- 24 Simmons, D., Satherthwaite, A.B., Tenen, D.G. & Seed, B. (1992) Molecular cloning of a cDNA encoding CD34, a sialomucin of human haematopoietic stem cells. *J. Immunol.* **148**, 267–271.
- 25 Osborn, L., Heission, C., Tizard, R. *et al.* (1989) Direct expression cloning of vascular cell adhesion molecule 1, a cytokine-induced endothelial protein that binds to lymphocytes. *Cell* **59**, 1203–1211.
- 26 Staunton, D., Dustin, M. & Springer, T. (1989) Functional cloning of ICAM-2, a cell adhesion ligand for LFA-1 homologous to ICAM-1. *Nature* **339**, 61–64.

.....

Chapter 19

Denaturing gradient gel electrophoresis and its variants in mutation detection

Rob B. van der Luijt & Riccardo Fodde*

Department of Human Genetics, Faculty of Medicine, Utrecht University, Stratenum, De Uithof Universiteitweg 100, 3584 CG Utrecht, The Netherlands

*MGC-Department of Human Genetics, Sylvius Laboratory, Faculty of Medicine, University of Leiden, Leiden, The Netherlands

.....

19.1	Introduction, 496	
19.1.1	Principles of DGGE, 496	
19.1.2	The introduction of a GC-clamp, 497	
19.1.3	Computational simulation of DNA melting behaviour, 498	
19.1.4	Perpendicular DGGE, 498	
19.1.5	Parallel DGGE, 499	
19.2	Examples of DGGE applications, 500	
19.3	General discussion, 500	
19.3.1	Applications of DGGE in molecular genetics, 500	
19.3.2	Advantages and disadvantages of the method, 500	
19.3.3	Variants of the DGGE method, 502	
19.3.3.1	Genomic denaturing gradient gel electrophoresis, 502	
19.3.3.2	Two-dimensional DNA typing, 502	
19.3.3.3	Constant denaturing gel electrophoresis, 502	
19.3.4	Comparison of DGGE with other techniques, 502	
19.3.4.1	DGGE vs. SSCP, 503	
	Protocol 99	
	Denaturing gradient gel electrophoresis, 503	
	Troubleshooting, 507	
	Power supply reads 60 V but gel does not run, 507	
	Fragments did not migrate far enough into the gel, 507	
	Fragments have migrated too far into the gel, 507	
	Fuzziness of the bands, 508	
	Extensive smiling of the bands, 508	
	References, 508	

19.1 Introduction

The detection and characterization of variations in nucleic acid sequence form an important component of the molecular genetic analysis of genomes. The introduction of polymerase chain reaction (PCR), which allows specific *in vitro* amplification of a particular target DNA sequence [1], has greatly facilitated the development of powerful techniques to identify genetic alterations. Several currently available screening techniques take advantage of changes in the physical properties of DNA caused by alterations in the nucleotide sequence. DNA sequence changes may result in differences in the melting behaviour of double-stranded DNA fragments or, alternatively, may modify the secondary structure of single-stranded DNA. As a consequence, fragments containing sequence alterations may display an altered mobility on gel electrophoresis. The precise molecular nature of the variant is then determined by DNA sequence analysis of such fragments.

Although PCR followed by direct DNA sequencing can be used to screen for unknown sequence alterations, it is often more practical to first determine which of the PCR-amplified DNA fragments contains a putative sequence alteration. This is especially the case when large DNA segments are being analysed, because it reduces the sequencing efforts required to further characterize the alteration. The large segment can be divided into overlapping PCR-amplified fragments which are then individually analysed. Only amplified fragments displaying an altered electrophoretic mobility (compared to the wild-type fragment) need then be sequenced. Denaturing gradient gel electrophoresis (DGGE) [2–4], which uses the melting properties of the double helix, provides one of the most sensitive protocols for the identification of sequence alterations in DNA. Moreover, several DGGE variants have been developed which are now widely applied to the molecular genetic analysis of genomic DNA.

Denaturing gradient gel electrophoresis (DGGE) is used to:

- identify germline and somatic mutations in genes
- analyse polymorphisms in genetic linkage, population and evolutionary studies
- establish mutational spectra induced by mutagens *in vitro* and *in vivo*
- examine the fidelity of DNA polymerases used in PCR

19.1.1 Principles of DGGE

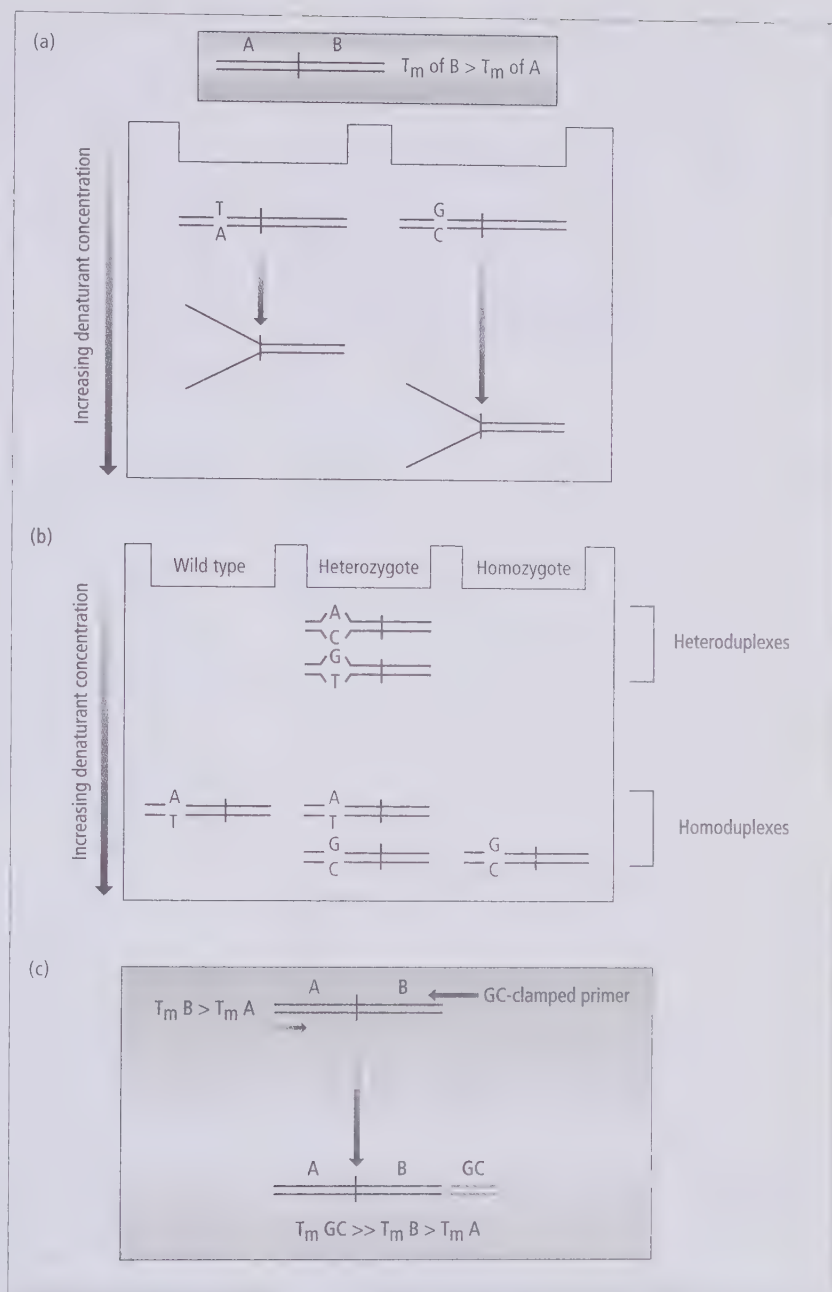
Denaturing gradient gel electrophoresis allows the resolution of DNA fragments differing by as little as a single nucleotide. The method is based on the differential electrophoretic mobilities of double-stranded (ds) wild-type and mutant DNA fragments through a linear gradient of increasing concentration of a denaturing agent (urea and formamide). The denaturing gradient can also be generated by temperature; this method is termed temperature gradient gel electrophoresis (TGGE).

The melting temperature (T_m) of a dsDNA molecule is defined as the temperature at which each base pair of the DNA duplex is in perfect equilibrium between the helical and denatured state. Within a DNA fragment, discrete regions with different T_m values (the so-called melting domains) may exist. The typical length of melting domains is generally between 50 and 300 bp under the conditions prevailing in denaturing gradient gels. As the DNA molecule migrates through the gel, the different melting domains will progressively denature into regions of single-stranded DNA. The latter phenomenon is dependent on the stability of the double helix, which in turn is determined by sequence composition (GC content) and by stacking interactions between adjacent basepairs. Therefore, the melting behaviour of a dsDNA molecule can be regarded as a strictly sequence-dependent property.

If a DNA fragment reaches a position along the denaturing gradient that equals the T_m of its lowest melting domain, partial denaturation or branching will occur. As a consequence, the electrophoretic mobility of the fragment will decrease markedly. DNA fragments differing by a single nucleotide change in their lowest melting domain can be separated on denaturing gradient gels because branching and consequent retardation of their mobility will occur at different positions along the gel (Fig. 19.1a). The melting temperature of a sequence in which an AT base pair is replaced by a GC base pair will slightly increase. Compared to the wild-type (AT) fragment, denaturation and mobility retardation of the mutant (GC) fragment will occur at a higher concentration of denaturing agent. The mutant fragment will migrate further through the gel before it reaches the position in the gradient that corresponds to the T_m of its lower melting domain. As a result, the wild-type and mutant DNA fragments will focus at different positions along the gel [5].

The resolving power of a denaturing gradient gel (i.e. the magnitude of separation between wild-type and mutant DNA fragments) is dependent on the

Fig. 19.1 DGGE analysis. (a) Schematic representation of the behaviour of two double-stranded DNA molecules, differing by a single nucleotide in their lowest melting domain (A) when analysed by DGGE. Partial denaturation (branching) and subsequent mobility retardation will occur at different positions along the gradient, resulting in separation of the two DNA fragments. (b) Schematic representation of DGGE analysis of a DNA sample heterozygous for a single base substitution. During PCR amplification of the target sequence, two homoduplexes and two heteroduplexes are generated, which are resolved on the DGGE gel. (c) Introduction of a GC-rich domain, the so called 'GC-clamp', prevents complete denaturation of the DNA fragment, allowing its analysis on DGGE. The target sequence is amplified by PCR. Because the 5' end of one of the members of the primer pair has a GC-rich extension (40 bp), the GC-clamp is added to the DNA fragment during PCR.



steepness of the gradient and gel running time. The sensitivity of DGGE is greatly enhanced when it is employed for the analysis of heterozygous nucleotide variants. During PCR amplification of the target sequence, the continuous denaturation and reannealing of DNA strands results in the formation of two homoduplexes as well as two heteroduplex molecules. Whereas a homoduplex consists of either two wild-type or two variant DNA strands, a heteroduplex is made up of a normal and a variant strand. The presence of a mismatch within a dsDNA molecule greatly decreases its melting

temperature, which causes the heteroduplexes to migrate more slowly than the two homoduplex molecules. This results in the appearance of two additional bands on the denaturing gradient gel which facilitates the visual detection of mutants (Fig. 19.1b).

19.1.2 The introduction of a GC-clamp

As mentioned above, DNA fragments usually consist of multiple melting domains which, upon migration of the DNA molecule through a dena-

turing gradient, undergo strand dissociation. In DGGE, single base changes in the lowest melting domain of a fragment will lead to differences in the pattern of electrophoresis in the denaturing gradient gel. DNA fragments with a nucleotide substitution within the highest T_m domain cannot be resolved due to the loss of sequence dependent migration upon complete strand dissociation. This problem, which initially limited the sensitivity of the DGGE procedure, has been circumvented by the attachment of a highly thermostable, GC-rich domain to the target sequence. Such a very high T_m domain (GC-clamp) prevents the target molecule from complete denaturation and allows the detection of variants in all melting domains [6]. A GC-clamp can efficiently be introduced during PCR amplification of the target sequence (Fig. 19.1c). By modifying one of the two amplification primers with a 5' GC-tail, the GC-rich domain will be incorporated during PCR at one of the ends of the resulting product. For the DGGE analysis of most DNA fragments, the 40-bp GC-clamp described by Sheffield *et al.* [7] can efficiently serve as high T_m domain. The nucleotide sequence of the GC-clamp is as follows:

5' - CGC.CCG.CCG.CGC.CCC.GCG.CCC.GTC.CCG.
CCG.CCC.CCG.CCC.C-3'.

However, if unusually GC-rich sequences are being analysed, longer GC-clamps may have to be employed [8]. The introduction of the GC-clamp increases the percentage of mutations detectable by DGGE to close to 100% [7].

19.1.3 Computational simulation of DNA melting behaviour

The possibility of simulating the melting behaviour of any known DNA sequence by computational analysis prior to the actual analysis of the samples of interest represents a great advantage of DGGE. The MELT87 (or its successor, MELT95) and SQHTX computer programs, as described by Lerman and Silverstein [9], allow a preliminary examination of the melting map for any DNA fragment of which the nucleotide sequence is available, and also the determination of optimal experimental conditions and of the expected effects of any base change on the melting map. The MELT87 program allows the identification of the different melting domains within a DNA molecule and their specific T_m s.

The presence of the GC-clamp on either side of the DNA fragment has profound effects on the melt map and on the percentage of detectable changes [10]. Using the information provided by this program, the optimal position of the GC-clamp, either at the 5' or 3' end of the fragment, may be chosen (Fig. 19.2).

If a DNA fragment has two distinct melting domains, the GC-clamp is usually added so that it will flank the highest T_m domain. Preferably, PCR primers should be designed in such a fashion that a 50- to 500-bp fragment, encompassing one or two melting domains, is generated. The presence of three or more melting domains within a single fragment should be avoided, since this usually results in a decreased detection sensitivity, especially within the highest melting domain. Significant T_m differences between two melting domains should also be avoided since branching of the lowest melting domain might cause electrophoretic retardation of the molecule to such a degree that it will not reach the point where the second melting domain is also denatured. If this cannot be avoided, two different gradients can be used to maximize separations resulting from changes in the different domains. Alternatively, the PCR product can be digested with a restriction enzyme, allowing separate analysis of the two melting domains. The SQHTX software program reports the expected difference in gradient level for a single base mismatch at every position along a fragment as a function of electrophoresis time. This program may be used to determine the optimal range of the denaturing gradient and the gel running time.

19.1.4 Perpendicular DGGE

The melting behaviour of a DNA molecule may also be determined experimentally by means of perpendicular DGGE. This approach is especially useful if the complete sequence of the fragment of interest is not known (to design primers, only the nucleotide sequences of the 5' and 3' ends are necessary). In a perpendicular DGGE, the denaturing gradient is perpendicular to the electrophoresis direction and the sample is applied along the entire width of the gel. Each DNA molecule will migrate through a constant denaturant concentration at a constant electrophoretic rate. The resulting curve is indicative of the number of melting domains and the percentage of the denaturing agents at which denaturation of each domain occurs. The latter allow the estimation of the denaturant concentration range to be used in parallel gels for the analysis of the same fragment. The preparation of perpendicular gels has been described in detail by Myers *et al.* [5] and will not be included here. However, the protocol reported here for parallel gels (Protocol 99) may also be applied to the casting of perpendicular ones provided that the denaturing gradient is poured perpendicular to the direction of electrophoresis.

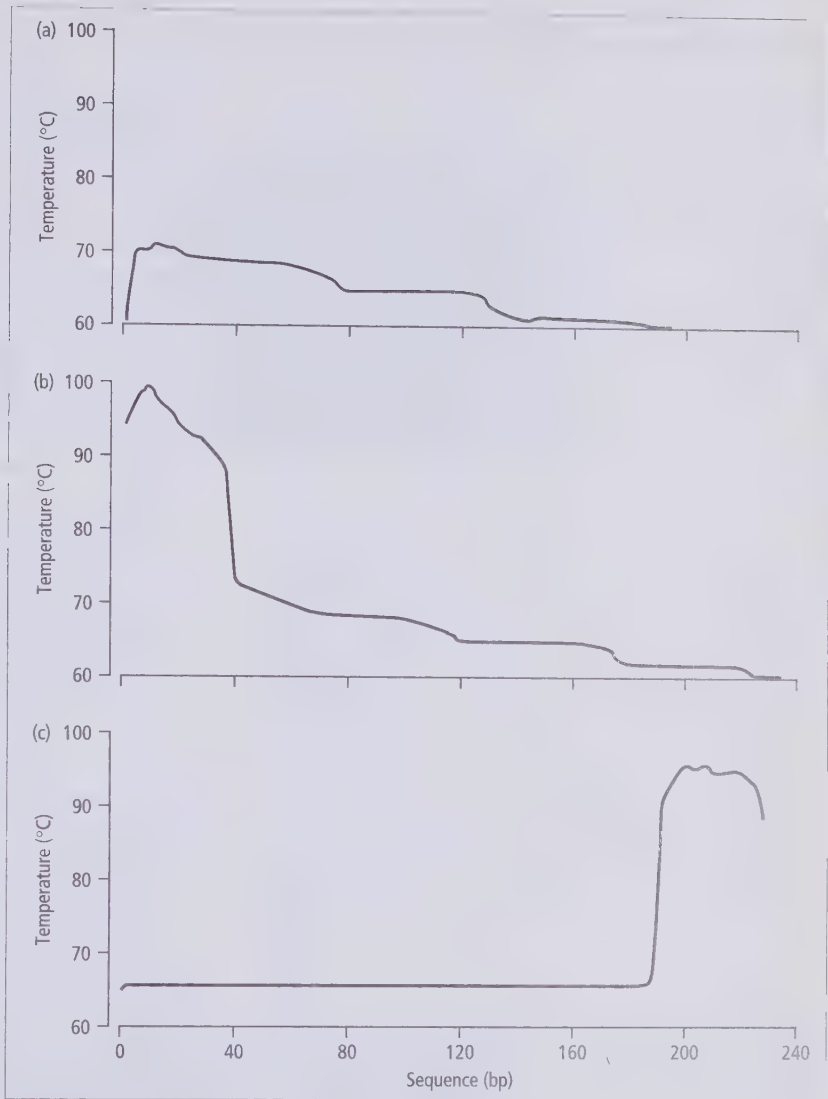


Fig. 19.2 Modification of the meltmap of a DNA fragment by attachment of a GC-clamp either at the 5' or 3' end. (a) Native meltmap of exon 4 of the human adenomatous polyposis coli (APC) gene (194 bp). (b) Meltmap of the same fragment after attachment of a 40-bp GC-clamp at the 5' end. The presence of three distinct melting domains might compromise mutation detection in the two highest domains. (c) Meltmap of the same fragment after attachment of the GC-clamp at the 3' end. The presence of a single melting domain makes the attachment of the GC-clamp at this end of the PCR product preferable.

19.1.5 Parallel DGGE

To detect DNA sequence alterations by DGGE, the PCR-amplified target sequence is subjected to electrophoresis through polyacrylamide gels containing a linear gradient of increasing concentrations of denaturing agents. The direction of the denaturing gradient in the gel is parallel to the path followed by the molecules during electrophoresis. The fragment migrates through the gel until it reaches a denaturant concentration where its mobility is abruptly retarded. As a result of this 'focusing', the sample will appear as a sharp band at a characteristic position along the gradient gel. The choice of the denaturant range is made based on the T_m of the domain of interest. Initially, parallel gels should be used with a top to bottom difference of 25–30% denaturant centred around the T_m of the domain. This usually results in a good resolution of

the wild type and mutant fragments. The conversion factor between T_m and percentage denaturant for gels run at 60 °C is given by the empirical formula:

$$\% \text{ denaturant} = (3.2 \times T_m) - 182.4$$

For example, if the melting domain of interest has a T_m of 75 °C (57.6% denaturant), then it should be ideally analysed on a parallel gel containing a range of 43–73% denaturant concentrations. When such a gradient is used, branching of the melting domain will occur when the fragment has run about half the distance of the gel. Narrower denaturant gradients (10–15% top to bottom difference) can also be employed with satisfactory results. If preliminary computational and/or experimental simulations have indicated the presence of two distinct melting domains within the same fragment, a different gradient should be designed for the optimal mutation analysis of each domain.

19.2 Examples of DGGE applications

In human molecular genetics, the DGGE method has been widely applied for the detection of mutations and polymorphisms in disease genes [11]. In our laboratory, this technique has been employed to perform mutation studies of genes involved in the pathogenesis of colorectal cancer. DGGE was used to identify somatic alterations of genes involved in the multistep process of colorectal tumorigenesis, such as *k-ras* (*KRAS2*) and *p53* (*TP53*) [12–14]. In patients with inherited colorectal cancer syndromes, DGGE was employed to screen for germline mutations at the *APC*, *MSH2* and *MLH1* genes [15–17]. As an example, we show here the application of the DGGE for the detection of germline *APC* mutations. Constitutional mutations of the tumour-suppressor gene *APC* are responsible for familial adenomatous polyposis (FAP), an autosomal dominantly inherited predisposition to colorectal cancer. Mutation studies of the *APC* gene are of importance for the presymptomatic diagnosis of the disease in predisposed individuals. Also, knowledge of the mutation spectrum of the *APC* gene may provide some insight in the mechanism of *APC*-driven tumorigenesis and may lead to the establishment of genotype–phenotype correlations. Several methods of mutation detection have been used to screen for *APC* mutations in patients with FAP, including the RNase protection assay [18,19], single-strand conformation polymorphism (SSCP) analysis [20,21] and DGGE [8,15,22,23]. These PCR-based methods permit the analysis of amplified segments of up to 500 bp in length. The first 14 exons, ranging in size from about 50 to 400 bp, are screened using an exon-by-exon strategy, while the unusually large exon 15 (± 6.5 kb) has to be divided into a large number (up to 23) of overlapping PCR-amplified fragments, which are separately analysed. Here we show examples of the results obtained by DGGE analysis of two *APC* exons (4 and 8).

19.3 General discussion

19.3.1 Applications of DGGE in molecular genetics

To date, the two main applications of the DGGE approach have been the direct detection of disease-causing mutations and the identification of polymorphisms in genomic sequences within or flanking ‘disease’ genes. In human molecular genetics, the PCR-DGGE protocol has been applied to the molecular analysis of a large number of genomic loci [11]. DGGE has proved a very rapid

and sensitive approach to the analysis of inherited conditions caused by heterogeneous mutation spectra or by frequent *de novo* mutations. Examples of such disorders include β -thalassaemia [24–27], haemophilia A and B [28–34], and cystic fibrosis [35]. In cancer genetics, the characterization of germline and somatic mutations, which lead to a malignant phenotype in a multistep process, has allowed the definition of important biological models such as the adenoma-carcinoma sequence in colorectal tumorigenesis [36]. In these cases, DGGE can be employed to detect alterations at tumour suppressor genes, oncogenes and DNA mismatch repair genes, such as *APC*, *TP53*, *KRAS2*, *MSH2* and *MLH1* [12–23] and monitor their accumulation in tumour progression.

More ‘research-oriented’ application of DGGE include the examination of the fidelity of several DNA polymerases [37], the analysis of the *in vitro* mutational spectra of several mutagens [38–44], evolutionary [45], and population genetic studies [46], and the detection of conformational transitions in nucleic acids [47].

19.3.2 Advantages and disadvantages of the method

Among the advantages of the technique are:

- 1 the high sensitivity of detection ($> 95\%$);
- 2 improved detection of heterozygotes (hetero-duplex formation);
- 3 the use of computer programs to optimize the analysis;
- 4 non-radioactive means of detection; and
- 5 easy isolation of the mutant allele for subsequent sequence determination.

Disadvantages of the DGGE can be represented by:

- 1 laborious and time-consuming preliminary work prior to the actual analysis of the fragment (computer or experimental simulations);
- 2 the costly synthesis of relatively long (≈ 60 nucleotides) PCR primers;
- 3 the limited size of the largest DNA fragment (≈ 500 bp) that can be efficiently analysed;
- 4 the use of special equipment; and
- 5 toxicity of some of the reagents (formamide).

While several alternatives to the use of GC-clamped primers have been successfully applied [48,49], DGGE analysis of a large genomic regions undeniably requires substantial preliminary work to maximize its efficiency. Nevertheless, the introduction of computer programs which allow simulation of the melting behaviour of DNA has considerably reduced the amount of preliminary experimental work needed.

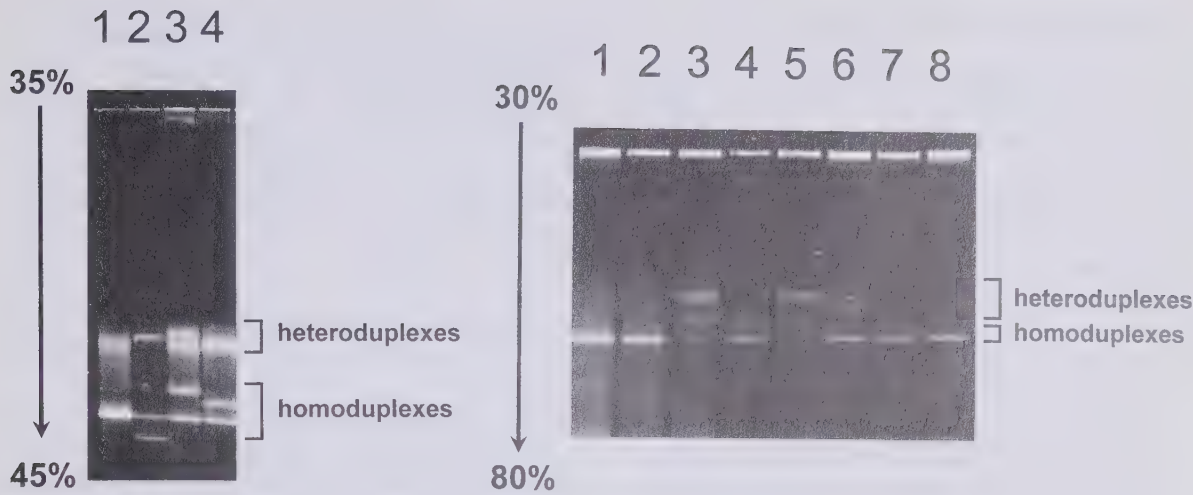


Fig. 19.3 Detection by DGGE of germ-line mutations of the APC gene in patients with familial adenomatous polyposis (FAP). (a) DGGE analysis of APC exon 4, amplified by PCR using genomic DNA samples obtained from a normal individual (lane 1) and unrelated FAP (lanes 2–4). The following germ-line APC mutations were identified by sequence analysis of the mutant homoduplexes: ATAG-deletion at codon 169–171 (lane 2), AGT → ATT substitution at codon 171 (lane 3), and C-deletion at codon 173 (lane 4). In the DGGE pattern of the fragment containing the single base change (lane 3), all four different molecular species (i.e. two homoduplexes and two heteroduplexes) can be distinguished. In the other two variant patterns, the heteroduplexes form a single band (the background smears in lanes 1, 3 and 4 are

caused by overloading). (b) DGGE analysis of APC exon 8, showing variant band pattern in four individuals. Lane 1 contains a normal control, lanes 2–8 are unrelated FAP patients. The samples in lanes 3 and 5 represent unrelated patients carrying the same mutation (a CT deletion at codon 298–299). Note that two distinct homoduplexes can be seen, but that the two heteroduplexes are not resolved. The exon 8 fragments in lanes 4 and 6 contain a CGA → TGA substitution at codon 302 (lane 4) and codon 283 (lane 6). Compared to the 2-bp deletion, the resolution of homo- and heteroduplexes in the case of the single base change is inverted. Whereas the heteroduplexes are now clearly separated, the homoduplexes have comigrated and appear as a single band.

Detection of APC gene mutations by DGGE

Perpendicular DGGE and computer simulation using the MELT87 program indicated that APC exon 4 and its intron–exon boundaries are encompassed within a single melting domain when the GC-clamp is positioned at the 3’ side of the 194 bp amplified fragment (see Fig. 19.2c). PCR amplification of exon 4 was performed using a primer pair of which one member contained the GC-clamp, while the other had a 23-bp universal M13 sequence to allow automated direct sequencing. Genomic DNA samples of unrelated FAP patients were amplified by PCR, and $\frac{1}{10}$ of the resulting product was analysed on 6% polyacrylamide gels containing a denaturing gradient ranging from 35 to 45%. Normal control samples showed a single, sharp band, corresponding to the exon sequence of the wild-type. DGGE analysis of the same fragment amplified from several of the FAP-affected individuals occasionally showed variant

band patterns, indicating the presence of sequence alterations within this exon (Fig. 19.3a). Nucleotide sequencing of the corresponding PCR products resulted in the identification of different pathogenic APC mutations. In patients with identical mutations, the band patterns on the DGGE gel are also identical. A typical example of the results obtained by DGGE analysis of another exon of the APC gene (exon 8) is shown in Fig. 19.3b. Compared with exon 4, where the top-to-bottom difference of the gradient was 10%, a much broader denaturing gradient was used for exon 8. For the DGGE analysis of this exon, a 50-bp GC-clamp [8] was added to the 3’ end of the target sequence. One tenth of the exon 8 PCR product was loaded on 6% polyacrylamide gels containing a 30–80% denaturing gradient. Several FAP patients showed variant DGGE patterns for this exon. Sequence analysis of the corresponding PCR products resulted in the identification of different germ-line mutations in this exon (Fig. 19.3b).

Case Study 19.1

19.3.3 Variants of the DGGE method

The DGGE technique was described for the first time in 1979 [2]. Since its introduction, the original protocol has been the object of many efforts to improve some of its features. In its basic and ideal form, GC-clamped DGGE will detect the great majority of base changes within a ≈ 500 -bp fragment encompassed within a single melting domain. However, when two or more T_m domains are present within the PCR fragment to be analysed, the resolution of mutations located within the most thermostable melting domain of the native molecule may become difficult. In order to maximize the efficiency of the DGGE-based strategy, several investigators have explored the possibility doing PCR on relatively large (2–3 kb) DNA targets and digesting the PCR product into ≈ 500 -bp fragments prior to DGGE [33]. Satoh *et al.* [34] implemented the same protocol by performing the PCR reactions with GC-clamped primers on both side of the target DNA fragment. The average theoretical detectability of this method should be ≈ 70 –75%, based on the 100% detectability for the two GC-clamped fragments and 50% for the intermediate ones.

19.3.3.1 Genomic denaturing gradient gel electrophoresis

In genomic DGGE (gDGGE), genomic DNA is digested with a restriction enzyme, electrophoresed through a denaturing gradient gel, transferred to nylon filters, and hybridized to a unique DNA probe [50]. Clear advantages of gDGGE over its parental protocol are:

- 1 it is not limited to any specific target sequence nor to its length (any available unique probe of any length can be used);
- 2 it does not require sequence information;
- 3 covalent modifications in genomic DNA otherwise lost by enzymatic amplification such as methylation are detectable.

On the other hand, because it relies on the presence of 'natural clamps' (a melting domain with a T_m higher than that of the domain where the putative variant is located) within the restriction fragments to be analysed, only a subset (20–60%) of all the possible single base variations will be detected by gDGGE (use of several restriction enzymes or combination of them might sensibly alleviate the latter problem). Moreover, since gDGGE is not PCR-based, heteroduplex formation is not feasible and one has to rely entirely on the resolution of two double-stranded DNA molecules differing by one base change. Nevertheless, gDGGE has been successfully applied for the identification

of polymorphic sequence variations in human chromosome 21 [51], and to screen for mutations in *Drosophila* [52–54]. Abrams *et al.* [55] have developed a modified protocol to generate heteroduplex molecules between a GC-clamped radiolabelled DNA probe and genomic DNA restriction fragments which are then analysed by DGGE. A similar approach is based on the DGGE analysis of heteroduplexes obtained by hybridization of radiolabelled RNA probes to either genomic restriction fragments [56] or GC-clamped PCR-amplified DNA [57].

19.3.3.2 Two-dimensional DNA typing

A two-dimensional fingerprint of complex genomes (two-dimensional DNA typing) [58] combines size fractionation of genomic restriction fragments in the first dimension with their sequence-dependent separation through denaturing gradient gels in the second dimension. Transfer to nylon membranes and hybridization with micro- and minisatellites or other repetitive sequences, results in complex patterns of spots (up to 600 depending on the probe), a portion of which are polymorphic among unrelated individuals. This technology has potentials for the analysis of genomic instability in relation to cancer, ageing and exposure to mutagen agents.

19.3.3.3 Constant denaturing gel electrophoresis

Another modification of the original DGGE protocol is constant denaturing gel electrophoresis (CDGE) [59]. Gels containing constant concentrations of the denaturing agent allow increased resolution of mutant fragments since they will constantly migrate with a different electrophoretic mobility throughout the whole length of the gel. Although CDGE has proven very useful for the screening of known mutations in the p53 [60] and *HPRT* genes [61], the method does not seem to represent a valid alternative to DGGE for the search of previously uncharacterized base changes in relatively large DNA fragments since each variant requires specific predetermined electrophoretic conditions for optimal resolution.

19.3.4 Comparison of DGGE with other techniques

For the detection of unknown sequence alterations in nucleic acids, different techniques with complementing strengths are currently available. The relative usefulness of these techniques depends upon a number of criteria, including sensitivity, reproducibility, rapidity and easiness. Only very few studies

have approached the systematic comparisons of the different available methods. When compared with RNase protection and chemical cleavage of mismatch using hydroxylamine and osmium tetroxide (HOT-CCM), DGGE proved more reliable and sensitive [57] combining a high sensitivity of mutation detection with a relatively less labour intensive protocol.

19.3.4.1 DGGE vs. SSCP

Due to its simplicity and relatively high sensitivity, the SSCP analysis is one of the most frequently used methods for the detection of sequence alterations. A protocol for SSCP can be found elsewhere in this volume (see Protocol 7, Chapter 5). A direct comparison between DGGE and SSCP has shown that the latter is less sensitive, since 90% of the variants could be detected compared with 100% in the case of DGGE [62]. Mutation studies of the APC gene in patients with familial adenomatous polyposis showed a much more dramatic difference between the detection rates of these two methods. Using GC-clamped DGGE, mutations were found in about 63% of the patients [8,22], which is comparable to results obtained in a similar study using the RNase protection assay. By the latter technique, mutations were found in 65% of the subjects [18,19]. In contrast, SSCP analysis revealed mutations in only 20% [20] and 29% [21] of the cases. In each of these studies, the entire APC coding sequence was investigated in a large series of patients. Although it should be noted that these studies were performed on different patient populations, it is unlikely that

this accounts for the marked difference in the observed detection rates.

Sheffield *et al.* [63] have demonstrated that the sensitivity of the SSCP is dependent on the length of the PCR product analysed. Previous authors showed that the sensitivity of mutation detection by SSCP drops dramatically with an increase in length of the fragment. Therefore, in order to allow efficient detection of sequence alterations by SSCP, PCR fragments should be kept relatively small. Whereas DGGE permits the analysis of fragments of about 500bp in length, the optimal fragment size for sensitive base substitution detection by SSCP is between 50 and 150bp. Although the decreased sensitivity of SSCP compared with DGGE can be improved by optimizing of the protocols employed, the consequential limited size of the PCR products represents a disadvantage of the SSCP method, especially when a large genomic region has to be analysed. Finally, apart from these intrinsic properties of the methods, the level of expertise for a certain procedure and the equipment available are important factors determining the relative usefulness of the above mentioned techniques.

In conclusion, DGGE has proved an efficient approach to the analysis of nucleic acids, and is employed for a wide spectrum of applications in both research and diagnostics. DGGE is a valid option when a very accurate and complete mutation analysis is required and when the time necessary to set up the technique and to define an optimal strategy do not compromise the speed at which the ultimate goal is achieved.

.....

Protocol 99 Denaturing gradient gel electrophoresis

For details of solutions, media and materials, see Appendix I. For suppliers and contact addresses see Appendix III.

DGGE is a conventional vertical polyacrylamide electrophoresis, where DNA molecules migrate through linearly increasing concentrations of denaturing agents (urea and formamide). Denaturing gradient gels are poured using conventional gradient makers. In order to ensure sharp, reproducible bands, constant temperature must be maintained within the gels. To maintain a uniform temperature in the gel, the plates enclosing the gel are submerged in a well-stirred, temperature-controlled bath of running buffer. The electrophoretic run is generally performed at 60 °C. This temperature was empirically chosen to exceed

the melting temperature of an AT-rich DNA fragment in the absence of denaturing agents. However, lower or higher bath temperatures can be used. For extremely GC-rich sequences, for example, temperatures up to 75 °C can be employed [64].

Materials

EQUIPMENT

The following list includes all the equipment required for the preparation and running of denaturing gradient gels.

- gel apparatus

The gel apparatus currently used in our laboratory is home-made, based on the original description by Myers *et al.* [5]. It consists of an acrylic frame suitable for holding the glass plates and gel submerged in a bath of the anode electrolyte maintained at 60 °C. Complete equipment kits are also commercially available from different companies (D-gene, Bio-Rad Laboratories, Hercules CA, USA; IngenyPhorU, Ingeny BV, Leiden, The Netherlands; BBS Scientific Company, Del Mar CA, USA). Alternatively, pre-existing vertical electrophoresis equipment (Protean II, Bio-Rad Laboratories; SE 600 Series, Hoefer Scientific Instruments, San Francisco CA, USA) can be adapted.

- two glass plates, one eared and one non-eared

The dimensions of the glass plates used in combination with the acrylic gel holder [5] are 18 cm wide × 20 cm high × 0.6 cm thick. The eared glass plate has a cutout 2 cm deep and 15 cm wide across the top.

- spacers and combs (Teflon, 0.6 mm thick)
- binder clips
- glass or acrylic aquarium

We use an aquarium tank 25 cm wide × 36 cm deep × 27 cm tall, which can be used to run two gels simultaneously. We usually perform DGGE in a volume of 14 litres running buffer.

- cathode (platinum)
- anode (platinum)
- combined thermostat and pump with tubing
- gradient maker (15–25 ml capacity per side)

A conventional gradient maker, composed of two cylindric reservoirs connected by a short tube at the base, is used to pour the denaturing gradient gel.

- power supply

REAGENTS AND SOLUTIONS FOR DGGE

The following list includes all the chemicals and solutions necessary to prepare and run denaturing gradient gels.

- 40% (w/v) acrylamide stock solution (acrylamide/*bis*-acrylamide, 37.5:1). Dissolve 100 g acrylamide and 2.7 g bisacrylamide in H₂O to a final volume of 250 ml. Store in dark glass bottles at 4 °C.

- 20× concentrated TAE electrophoresis buffer: 0.8 M Tris base, 0.4 M sodium acetate, 0.02 M EDTA (pH 8.0).
Dissolve 97 g Tris base, 7.5 g Na₂EDTA and 54.5 g sodium acetate 3H₂O in H₂O to 1000 ml. Adjust pH to 8.0 with glacial acetic acid (≈ 36 ml). Store in dark glass bottles at 4 °C.
- 6%^a (w/v) acrylamide stock solution (0% denaturant stock solution) in TAE buffer: for 500 ml: 75 ml acrylamide (40% stock), 25 ml TAE (20× stock), and H₂O up to 500 ml.
- 80% denaturant stock solution (6%^a acrylamide, 32% formamide, 5.6 M urea): for 500 ml: 170 g electrophoresis-grade urea, 75 ml acrylamide (40% stock), 160 ml deionized formamide^b (100% stock), 25 ml TAE buffer (20× stock), and H₂O to 500 ml. Store in dark glass bottles at 4 °C.
- 10% (w/v) ammonium persulphate stock solution. Dissolve 10 g ammonium persulphate to 100 ml H₂O. This solution is usually freshly prepared but may also be stored in small aliquots (1 ml) at –20 °C.
- TEMED (*N,N,N',N'*-tetramethylethylenediamine).
- 5× gel loading solution (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 20% (w/v) Ficoll). Dissolve 20 g Ficoll, 250 mg bromophenolblue and 250 mg xylene cyanol in a final volume of 100 ml H₂O. Instead of Ficoll, glycerol can be used.
- 10 mg ml⁻¹ ethidium bromide: dissolve 1 g ethidium bromide in 100 ml H₂O.

^a Different acrylamide percentages may be employed depending on DNA fragment size.

^b To deionize formamide: add 2 g of mixed bed resin (Baker) to 100 ml formamide and stir for 30 min. Filter to remove resin and store in dark glass bottles.

Method

PREPARATION OF DENATURING GRADIENT GELS

Parallel gels contain a concentration gradient of formamide and urea linearly increasing from the top to the bottom of the gel. The gels are used to analyse a large number (20–30) of samples, which are loaded into wells at the top of the gel.

- 1 Thoroughly clean the glass plates, spacers, comb and gradient maker with a strong detergent. Rinse the plates with ethanol and dry them carefully.
- 2 Arrange the spacers along the sides of the larger (non-eared) plate. Lay the eared plate in position and clamp them together with binder clips. Carefully seal the sides and bottom with gel-sealing tape. Spacers may be greased lightly to prevent leakage.
- 3 Put the comb into position and place the glass plates in the gel frame so that the eared glass plate is facing the rubber gasket, thus forming the upper electrophoresis chamber. Slide the acrylic braces between the outer non-eared glass plate and the tightening screws. Clamp the glass plates into position by tightening the screws. Leave room for air to escape.
- 4 Place the gradient maker on top of a magnetic stirrer, about 25 cm above the gel frame. Insert the exit-tube of the gradient maker

between the glass plates next to the comb. Make sure that this tube as well as the tube connecting the two chambers of the gradient makers are closed. Prepare two solutions of equal volume (15 ml; in our set-up a 30 ml volume will just fill the plates) which will give the desired denaturant concentration range. Add 16 μ l TEMED and 160 μ l 10% APS to each solution and mix well.

- 5** Pour 15 ml of the solution with the higher concentration of denaturant in the chamber of the gradient maker that is connected to the plate cavity. Briefly open and close the connection between the chambers so to allow the solution to fill the connecting tube. Make sure that no air bubbles block the passageway between the two chambers.
- 6** Pour 15 ml of the solution with the lower percentage of denaturant in the other chamber.
- 7** While stirring both solutions, open the connection between the two chambers and the exit-tube to the glass plates. Avoid air bubbles.
- 8** The liquid passes by gravity through the plastic tubing into the cavity between the two glass plates. The gel should take about 5 min to pour in.
- 9** Allow the gel to polymerize for about 30–60 min.
- 10** Gently remove the comb from the gel and the tape from the bottom of the glass plates.

GEL ELECTROPHORESIS

- 11** Place the frame with the gel into the bath containing the 1 \times TAE buffer heated to 60 °C. Adjust the volume of the buffer so that it just rises above the level of the wells. Avoid contact of the buffer with the upper electrophoresis chamber. Connect the tubing of the combination thermostat so that buffer circulates from the aquarium into the upper buffer chamber (containing the cathode), while it overflows through a hole in the rear of the frame into the aquarium.
- 12** Pre-run the gel for about 30 min at 60 V (about 50 mA).
- 13** Add gel loading solution to the samples. Depending on the yield of the PCR reaction, we usually load between one-tenth and a half of the total PCR product. A small final volume (\pm 10 μ l) will result in sharper bands.
- 14** With a syringe fitted with a needle, flush the wells with 1 \times TAE buffer. Load the samples and start the electrophoresis. For reasons of convenience, we usually perform our DGGE runs for about 16 h at 60 V (50 mA); however, different times and running conditions can be applied.

STAINING THE GEL

- 15** Stop the electrophoresis and remove the gel frame from the aquarium tank. Remove the glass plates from the frame and gently lift the eared glass plate; use the other plate to support the gel during staining.
- 16** Stain the gel for 20–30 min in 250 ml 1×TAE containing 0.5 µg ml⁻¹ ethidium bromide with gentle shaking.
- 17** If a high background is observed, a destaining step of 15–20 min in 250 ml 1×TAE (or water) can be introduced.
- 18** Examine the gel under UV (254 nm).

Troubleshooting

Power supply reads 60 V but gel does not run

This is usually caused by a short circuit somewhere in the electrophoresis assembly.

- *Check whether the electrodes are correctly connected to the power supply (they may have been switched). The cathode should be placed in the upper buffer chamber of the gel frame. The anode should be in contact with the running buffer in the aquarium tank.*
- *Check if the sealing tape was removed from the bottom of the glass plates!*
- *Make sure the buffer is circulated properly through the upper electrophoresis chamber (it should overflow only through the hole in the rear of the buffer compartment).*

Fragments did not migrate far enough into the gel

When the concentration of denaturing agent at the top of the gel or the temperature of the running buffer are too high, DNA fragments will start denaturing prematurely. As a consequence, these fragments will not migrate further into the gel and will not reach the position where the resolving power of the denaturing gradient gel is optimal.

- *Check the temperature of the buffer (it may be too high).*
- *Check the appropriateness of the denaturing gradient.*
- *Adjust the gradient by lowering the concentration of denaturant at the top of the gel.*

Fragments have migrated too far into the gel

- *Check the temperature of the buffer (it may be too low).*
- *Check the appropriateness of the denaturing gradient. Adjust the gradient by increasing the concentration of denaturant at the at the bottom of the gel.*

Fuzziness of the bands

Fuzzy bands or smears are usually visible when the DNA fragments have not properly focused or when the gel is overloaded.

- *Run the gel for a longer period.*
- *Load a smaller amount of sample on the gel.*
- *Prepare fresh gel solutions (denaturing agents may have decayed).*

In some cases, fuzzy bands or smears are the result of improper melting behaviour of the DNA fragment due to an inadequate choice of DGGE primers and/or denaturing gradient.

Extensive smiling of the bands

Smiling is often caused by a leakage current passing through a channel between the polyacrylamide gel and the spacers. Such channels are formed by shrinkage of polyacrylamide gels upon polymerization. By gently pushing the spacers against the gel after the gel has polymerized, this gel running artefact may be prevented.

References

- 1 Saiki, R.K., Gelfand, D.H., Stoffel, S. *et al.* (1988) Primer directed amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491.
- 2 Fischer, S.G. & Lerman, L.S. (1979) Length-independent separation of DNA restriction fragments in two-dimensional gel electrophoresis. *Cell* **16**, 191–200.
- 3 Fischer, S.G. & Lerman, L.S. (1983) DNA fragments differing by a single base-pair substitution are separated in denaturing gradient gels: correspondence with melting theory. *Proc. Natl Acad. Sci. USA* **80**, 1579–1583.
- 4 Myers, R.M., Lumelski, N., Lerman, L.S. & Maniatis, T. (1985) Detection of single base substitutions in total genomic DNA. *Nature* **313**, 495–498.
- 5 Myers, R.M., Maniatis, T. & Lerman, L.S. (1987) Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Meth. Enzymol.* **155**, 501–527.
- 6 Myers, R.M., Fischer, S.G., Lerman, L.S. & Maniatis, T. (1985) Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Res.* **13**, 3131–3145.
- 7 Sheffield, V.C., Cox, D.R., Lerman, L.S. & Myers, R.M. (1989) Attachment of a 40 base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single base changes. *Proc. Natl Acad. Sci. USA* **86**, 232–236.
- 8 Olschwang, S., Laurent-Puig, P., Groden, J., White, R. & Thomas, G. (1993) Germ-line mutations in the first 14 exons of the adenomatous polyposis coli (APC) gene. *Am. J. Hum. Genet.* **52**, 273–279.
- 9 Lerman, L.S. & Silverstein, K. (1987) Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. *Meth. Enzymol.* **155**, 482–501.
- 10 Myers, R.M., Fischer, S.G., Maniatis, T. & Lerman, L.S. (1985) Modification of the melting properties of duplex DNA by attachment of a GC-rich DNA sequence as determined by denaturing gradient gel electrophoresis. *Nucleic Acids Res.* **13**, 3111–3130.
- 11 Fodde, R. & Losekoot, M. (1994) Mutation detection by denaturing gradient gel electrophoresis (DGGE). *Hum. Mut.* **3**, 83–94.
- 12 Pellegata, N.S., Losekoot, M., Fodde, R. *et al.* (1992) Detection of k-ras mutations by denaturing gradient gel electrophoresis (DGGE): a study on pancreatic cancer. *Anticancer Res.* **12**, 1731–1736.
- 13 Renault, B., van den Broek, M., Fodde, R. *et al.* (1993) Base transitions are the most frequent genetic changes at p53 in gastric cancer. *Cancer Res.* **53**, 2614–2617.
- 14 van den Broek, M.H., Renault, B., Fodde, R. *et al.* (1993) Sites and types of p53 mutations in an unselected series of colorectal cancers in the Netherlands. *Anticancer Res.* **13**, 587–592.
- 15 Fodde, R., van der Luijt, R., Wijnen, J.Th., *et al.* (1992) Eight novel inactivating germ line mutations identified by denaturing gradient gel electrophoresis. *Genomics* **13**, 1162–1168.
- 16 Wijnen, J.Th., Vasen, H., Meera Khan, P. *et al.* (1995) Seven new mutations in hMSH2, a HNPCC gene, identified by denaturing gradient gel electrophoresis. *Am. J. Hum. Genet.* **56**, 1060–1066.
- 17 Wijnen, J.Th., Meera Khan, P., Vasen, H. *et al.* (1996) Majority of hMLH1 mutations responsible for hereditary nonpolyposis colorectal cancer cluster at the exonic region 15–16. *Am. J. Hum. Genet.* **58**, 300–307.

- 18 Miyoshi, Y., Ando, H., Nagase, H. *et al.* (1992) Germ-line mutations of the APC gene in 53 familial adenomatous polyposis patients. *Proc. Natl Acad. Sci. USA* **89**, 4452–4456.
- 19 Nagase, H., Miyoshi, Y., Horii, A. *et al.* (1992) Screening for germ-line mutations in familial adenomatous polyposis patients: 61 new patients and a summary of 150 unrelated patients. *Hum. Mut.* **1**, 467–473.
- 20 Groden, J., Gelbert, L., Thliveris, A. *et al.* (1993) Mutational analysis of patients with adenomatous polyposis: identical inactivating mutations in unrelated individuals. *Am. J. Hum. Genet.* **52**, 263–272.
- 21 Varesco, L., Gismondi, V., James, R. *et al.* (1993) Identification of APC gene mutations in italian adenomatous polyposis coli patients by PCR-SSCP analysis. *Am. J. Hum. Genet.* **52**, 280–285.
- 22 Olschwang, S., Tiret, A., Laurent-Puig, P. *et al.* (1993) Restriction of ocular fundus lesions to a specific subgroup of APC mutations in adenomatous polyposis coli patients. *Cell* **75**, 959–968.
- 23 Wallis, Y.L., MacDonald, F., Hultén, M. *et al.* (1994) Genotype–phenotype correlation between position of constitutional APC gene mutation and CHRPE expression in familial adenomatous polyposis. *Hum. Genet.* **94**, 543–548.
- 24 Cai, S.P. & Kan, Y.W. (1990) Identification of the multiple β -thalassemia mutations by denaturing gradient gel electrophoresis. *J. Clin. Invest.* **85**, 550–553.
- 25 Losekoot, M., Fodde, R., Harteveld, C.L. *et al.* (1990) Denaturing gradient gel electrophoresis and direct sequencing of PCR amplified genomic DNA: a rapid and reliable diagnostic approach to β -thalassemia, *Br. J. Haematol.* **76**, 269–274.
- 26 Losekoot, M., Fodde, R., Harteveld, C.L. *et al.* (1991) Homozygous β^0 -thalassaemia owing to a mutation in the cleavage-polyadenylation sequence of the human β -globin gene. *J. Med. Genet.* **28**, 252–255.
- 27 Rosatelli, M.C., Dozy, A., Faa, V. (1992) Molecular characterization of β -thalassemia in the Sardinian population. *Am. J. Hum. Genet.* **50**, 422–426.
- 28 Kogan, S. & Gitschier, J. (1990) Mutations and polymorphisms in the Factor VIII gene discovered by denaturing gradient gel electrophoresis. *Proc. Natl Acad. Sci. USA* **87**, 2092–2096.
- 29 Traystman, M.D., Higuchi, M., Kasper, C.K. *et al.* (1990) Use of denaturing gradient gel electrophoresis to detect point mutations in the factor VIII gene. *Genomics* **6**, 293–301.
- 30 Higuchi, M., Kazazian, H.H. Jr, Kasch, L. *et al.* (1991) Molecular characterization of severe haemophilia A suggests that about half the mutations are not within the coding regions and splice junctions of the factor VIII gene. *Proc. Natl Acad. Sci. USA* **88**, 7405–7409.
- 31 Higuchi, M., Antonarakis, S.E., Kasch, L. *et al.* (1991) Molecular characterization of mild-to-moderate haemophilia A: detection of the mutation in 25 of 29 patients by denaturing gradient gel electrophoresis. *Proc. Natl Acad. Sci. USA* **88**, 8307–8311.
- 32 Aly, A.M., Higuchi, M., Kasper, C.K. *et al.* (1992) Haemophilia A due to mutations that create new N-glycosylation sites. *Proc. Natl Acad. Sci. USA* **89**, 4933–4937.
- 33 Attree, O., Vidaud, D., Vidaud, M. *et al.* (1989) Mutations in the catalytic domain of human coagulation factor IX: rapid characterization by direct genomic sequencing of DNA fragments displaying an altered melting behavior. *Genomics* **4**, 266–272.
- 34 Satoh, C., Takahashi, N., Asakawa, J. *et al.* (1993) Variations among Japanese of the factor IX gene (F9) detected by PCR-denaturing gradient gel electrophoresis. *Am. J. Hum. Genet.* **52**, 167–175.
- 35 Ferec, C., Audrezet, M.P., Mercier, B. *et al.* (1992) Detection of over 98% cystic fibrosis mutations in a Celtic population. *Nature Genet.* **1**, 188–191.
- 36 Fearon, E.R. & Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767.
- 37 Keohavong, P. & Thilly, W.G. (1989) Fidelity of DNA polymerase in DNA amplification. *Proc. Natl Acad. Sci. USA* **86**, 9253–9257.
- 38 Cariello, N.F., Swenberg, J.A., De Bellis, A. & Skopek, T.R. (1991) Analysis of mutations using PCR and denaturing gradient gel electrophoresis. *Env. Mol. Mutagen.* **18**, 249–254.
- 39 Thilly, W.G. (1985) Potential use of denaturing gradient gel electrophoresis in obtaining mutational spectra from human cells. In *Carcinogenesis: The Role of Chemicals and Radiation in the Etiology of Cancer* (Huberman, E. & Barr, S.H., eds) vol. 10, 511–528 (Raven Press, New York).
- 40 Cariello, N.F. & Thilly, W.G. (1986) Use of denaturing gradient gel electrophoresis to determine mutational spectra in human cells. In *Mechanisms of DNA damage and Repair: Implications for Carcinogenesis and Risk Assessment* (Simic, M.G. *et al.*, eds) 439–452 (Plenum Press, New York).
- 41 Cariello, N.F., Keohavong, P., Kat, A.G. & Thilly, W.G. (1990) Molecular analysis of complex human cell populations: mutational spectra of MNNG and ICR-1191. *Mutat. Res.* **231**, 165–176.
- 42 Oller, A.R. & Thilly, W.G. (1992) Mutational spectra in human B-cells. Spontaneous, oxygen and hydrogen peroxide-induced mutations at the hprt gene. *J. Mol. Biol.* **228**, 813–826.
- 43 Keohavong, P. & Thilly, W.G. (1992) Mutational spectrometry: a general approach for hot-spot point mutations in selectable genes. *Proc. Natl Acad. Sci. USA* **89**, 4623–4627.
- 44 Chen, J. & Thilly, W.G. (1994) Mutational spectrum of chromium (VI) in human cells. *Mutat. Res.* **323**, 21–27.
- 45 Rogers, J., Ruano, G. & Kidd, K.K. (1992) Variability in nuclear DNA among nonhuman primates: application of molecular genetic techniques to intra- and inter-species genetic analysis. *Am. J. Primatol.* **27**, 93.
- 46 Takahashi, N., Hiyama, K., Kodaira, M. & Satoh, C. (1991) The length polymorphism in the 5' flanking region of the human β -globin gene with denaturing gradient gel electrophoresis in a Japanese population. *Hum. Genet.* **87**, 219–220.
- 47 Weber, C.K., Shaffer, D.J. & Sidman, C.L. (1991) Unexpected behaviour of H2Kb mutant DNAs in denaturing gradient gel electrophoresis. *Nucleic Acids Res.* **19**, 3331–3335.
- 48 Top, B. (1992) A simple method to attach a universal 50-bp GC-clamp to PCR fragments used for mutation analysis by DGGE. *PCR Meth. Appl.* **2**, 83–85.

- 49 Costes, B., Girodon, E., Ghanem, N. *et al.* (1993) Psoralen-modified oligonucleotide primers improve detection of mutations by denaturing gradient gel electrophoresis and provide an alternative to GC-clamping. *Hum. Mol. Genet.* **2**, 393–397.
- 50 Borresen, A.L., Hovig, E. & Brogger, A. (1988) Detection of base mutations in genomic DNA using denaturing gradient gel electrophoresis (DGGE) followed by transfer and hybridization with gene-specific probes. *Mutat. Res.* **202**, 77–83.
- 51 Burmeister, M., diSibio, G., Cox, D.R. & Myers, R.M. (1991) Identification of polymorphisms by genomic denaturing gradient gel electrophoresis: application to the proximal region of human chromosome 21. *Nucleic Acids Res.* **19**, 1475–1481.
- 52 Doerig, R.E., Suter, B., Gray, M. & Kubli, E. (1988) Identification of an amber nonsense mutation in the *rosy516* gene by germline transformation of an amber suppressor tRNA gene. *EMBO J.* **7**, 2579–2584.
- 53 Curtis, D., Clark, S.H., Chovnick, A. & Bender, W. (1989) Molecular analysis of recombination events in *Drosophila*. *Genetics* **122**, 653–661.
- 54 Gray, M., Charpentier, A., Walsh, K. *et al.* (1991) Mapping point mutations in the *Drosophila rosy* locus using denaturing gradient gel blots. *Genetics* **127**, 139–149.
- 55 Abrams, E.S., Murdaugh, S.E. & Lerman, L.S. (1990) Comprehensive detection of single base changes in human genomic DNA using denaturing gradient gel electrophoresis and a GC clamp. *Genomics* **7**, 463–475.
- 56 Takahashi, N., Hiyama, K., Kodaira, M. & Satoh, C. (1990) An improved method for the detection of genetic variations in DNA with denaturing gradient gel electrophoresis. *Mutat. Res.* **234**, 61–70.
- 57 Theophilus, B.D.M., Latham, T., Grabowski, G.A. & Smith, F.I. (1989) Comparison of RNase A, chemical cleavage, and GC-clamped denaturing gradient gel electrophoresis for the detection of mutations in exon 9 of the human acid β -glucosidase gene. *Nucleic Acids Res.* **17**, 7707–7722.
- 58 Uitterlinden, A.G., Slagboom, P.E., Knook, D.L. & Vijg, J. (1989) Two dimensional DNA fingerprinting of human individuals. *Proc. Natl Acad. Sci. USA* **86**, 2742–2746.
- 59 Hovig, E., Smith-Sorensen, B., Brogger, A. & Borresen, A.L. (1991) Constant denaturant gel electrophoresis, a modification of denaturing gradient gel electrophoresis, in mutation detection. *Mutat. Res.* **262**, 63–71.
- 60 Borresen, A.L., Hovig, E., Smith-Sorensen, B. *et al.* (1991) Constant denaturant gel electrophoresis as a rapid screening technique for p53 mutations. *Proc. Natl Acad. Sci. USA* **88**, 8405–8409.
- 61 Smith-Sorensen, B., Hovig, E., Andersson, B. & Borresen, A.L. (1992) Screening for mutations in the human HPRT cDNA using the polymerase chain reaction (PCR) in combination with constant denaturant gel electrophoresis (CDGE). *Mutat. Res. Fundam. Mol. Mech. Mutagen.* **269**, 41–53.
- 62 Moyret, C., Theillet, C., Puig, P.L. *et al.* (1994) Relative efficiency of denaturing gradient gel electrophoresis and single strand conformation polymorphism in the detection of mutations in exon 5 to 8 of the p53 gene. *Oncogene* **9**, 1739–1743.
- 63 Sheffield, V.C., Beck, J.S., Kwitek, A.E. *et al.* (1993) The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics* **16**, 325–332.
- 64 Hartevelde, C.L., Heister, J.G.A.M., Giordano, P.C. *et al.* (1996) Rapid detection of point mutations and polymorphisms of the α -globin genes by DGGE and SSCA. *Hum. Mut.* **7**, 114–122.

Index

Page numbers in *italic type* refer to figures and tables; those in **bold** refer to protocols

abortion, recurrent 156
abscisic acid sensitivity control 773
ACEDB database 691, 692, 693, 812, 834, 835
(AC)_n repeats 108–9
acute leukaemia symptoms 164
acute lymphoblastic leukaemia 162
acute lymphocytic leukaemia 160
acute lymphoid leukaemia (ALL)
 chromosome changes 978–9
 classification 979
acute myeloid leukaemia (AML) 160, 162, 163, 164
 chromosome changes 977–8
 chromosome painting 248, 249
 classification 978
acute non-lymphocytic leukaemia (ANLL) 162
acute promyelocytic leukaemia (APML) 163
ADA gene 653–4
addresses 879–80
adenine phosphoribosyltransferase 333
adenine plus alanosine medium 333
adenomatous polyposis coli (APC) 82
adenosine deaminase (ADA) 653
 deficiency and gene therapy 662, 663
adenoviral vectors 662
adherent cell–adherent cell fusions 327
Aegilops speltoides 747
AFLP scanning 113
AGAMOUS gene 779

age, breast cancer liability classes 32
AIDS, gene therapy 652, 660–1
albizzin/asparagine synthetase 337
alkaline phosphatase
 chemiluminescent substrates 604
 fluorogenic substrates 606
 sequence labelling 603
alkaline phosphatase-labelled antibodies 603
allele
 drop-out 134
 frequencies 130
 marker 39–40
allele-specific oligonucleotide (ASO) 106
allelic heterogeneity 33
allopolyploid 746
alpha satellite probes 217
Alu consensus 620, 621
Alu elements 109, 111
 human 110
Alu paints 299
Alu repeat 244–5
 exon amplification 446
Alu sequences, repeat analysis 619, 620, 621
Alu subfamily identification 620, 622
Alu-PCR 108, 244–5
 banding 245
 chromosome paints 156, 249, 296
 generation 298–9
 flow sorting 294
 flow-sorted chromosomes
 abnormal Plate 5, 246
 amplification 252–3

 with primer AGK34 227–8
somatic cell hybrid characterization 345
whole chromosome painting probes
 Plate 4
YACs 218
 amplified sequences 249
alveolar rhabdomyosarcoma 194, 195
 genomic hybridization analysis Plate 1
 interphase FISH analysis Plate 2
Alzheimer's disease 656
7-amino-4-methylcoumarin-3-acetic acid (AMCA) 215
ammonium persulphate (APS) 581
amniocentesis 153
amniotic fluid
 cultures 173–4
 harvesting 153–4, 177–9
 prenatal diagnosis 153
 unsynchronized cultures 277
amphidiploids, plant genome analysis 750–1
amphiplasty 748
ampicillin resistance 375
amplification refractory mutation system (ARMS) 107
amplified length polymorphisms (AMPFLP) 134
 data interpretation 135
 anchored island 433, 434
 length 436
 number 436
 proportion of genome not covered by 436
 singleton 434, 435
annealing, probe ordering 427, 429

- anonymous ftp 815, 816, 820–33
Anopheles gambiae 683
 anti-BrdU antibodies 216
 anti-BrdU-FITC 235
Antirrhinum, *myb*-related genes 780
 antisense constructs 657
 APC gene 500, 501
 mutations 501, 503
 aplets 835
 apoptosis 656–7
 APRT gene 338
Arabidopsis Plate 10, 632, 762
 AGAMOUS gene 779
 cDNA sequences 811
 chloroplast genome 763
 chromosome number 887
 Cnx1 protein 780–1
 cross hybridization of genes 779
 ESSA project 770, 778
 generation time 762
 genetic maps 764–5
 amplified fragment length
 polymorphisms (AFLPs) 765
 expressed sequence tags (ESTs) 765
 molecular markers 765
 mutants 764
 recombinant inbred lines 765
 genome
 mapping project 438
 size 762–4, 888
 Information Management System
 (AIMS) 775, 778
 mitochondrial genome 763
 model for non-plant species 780–1
 myb-related genes 780
 physical map 765–6
 RFLP markers 766
 seed size 762
 software 775
 communication tools 777
 stock centres 774–5, 777
 useful addresses 777–8
 YAC clone markers 766
 YAC libraries 765–6
ARABIDOPSIS newsgroup 775–6, 777
Arabidopsis nuclear genome 762, 763
 AtDB database 775, 776
 Athila retrotransposon 763
 cDNA sequencing 386–90
 classical gene identification 766
 DNA renaturation kinetics 763
 enhancer trap 769
 expressed sequence tags (ESTs) 769–73
 gene
 density 773
 duplication 774
 identification strategies 766–9
 intron number 773
 genetic complexity 765
 genomic sequencing 773–4
 map-based cloning 766–9
 microsatellites 763
 model for other species 779–81
 multigene families 772, 773
 mutant complementation 766–7
 organization 763
 plasma membrane integral proteins 766
 promoter trapping 769
 regulatory protein 773
 repeat sequences 763
 sequencing of large portions 774
 size 763
 T-DNA 767–8, 769
 Tat1 element 763
 Tat1 763
 transposon tagging 768–9
 tumour suppressor gene homologues
 772
 asnA gene 337
 asparagine synthetase 337
 asynapsis 751
 ataxia telangiectasia 475
 AtDB database 764, 775, 776
 Atto-Phos 606
 autoallotriploid hybrids 750
 autopolyploid 746
 auxotrophic mutants 333–4
 avian myeloblastosis virus reverse
 transcriptase 565
 avidin-fluorescein isothiocyanate *see*
 fluorescein isothiocyanate (FITC)
 B-cell mitogen 166
 Bacillus megaterium genome size 888
 Bacillus subtilis genome sequence 710, 736
 back-crosses 48
 bacterial artificial chromosomes (BACs)
 371
 rice physical map 805
 vector system 642
 bacterial infection 326
 bacterial interspersed mosaic element
 (BIMES) 729
 bacteriophage *see* phage
 Bal31 523
 deletion strategy 560
 exonuclease 558
 *Bam*HI/*Bgl*II cosmid digest 445, 446
 barr program 430, 432, 433
 BC1 progeny 786
 BCR-ABL fusion gene 151
 BELL1 gene 780
 bioinformatics 811
 biotin 218–19, 220, 228–30
 detection with FITC 235
 detection with Texas red 235–6
 DNA labelling by nick translation
 228–30
 quality control of labelling 230–1
 sequence labelling 602, 603
 biotin-11-dUTP, PCR product labelling
 296
 biotin-streptavidin system 551–2
 bis-acrylamide 580, 581
 BLASTN program 624
 BLASTX program 771
 ble gene 337
 bleomycin- and phleomycin-binding
 protein 337
 blood bottle preparation 869
 blood cell transformation with
 Epstein-Barr virus 871
 blood culture, unsynchronized 276
 blood samples
 cryopreservation 275–6
 lymphocyte separation 274
 processing 168–9
 slide preparation 169–70
 blood tube processing 869–70
 Bloom's syndrome 475–6
 bobbed (bb) mutation 673
 bone marrow culture, unsynchronized 276
 bone marrow samples
 cryopreservation 275–6
 synchronization technique 184–5
Brassica genome 779
BRCA1 gene 28, 29, 34
 linkage 34
BRCA2 gene 34
 BrdU *see* 5-bromo-2-deoxyuridine (BrdU)
 breast, phyllodes tumour 192
 breast cancer
 A allele 32
 age 32
 family pedigrees 29, 30
 G-banding in carcinoma 192
 inherited susceptibility 28
 liability classes 31, 32
 linkage analysis 30–3
 linkage heterogeneity 34–5
 media for cytogenetics 196
 recombination fraction 31
 sarcoma 196
 segregation analyses 30
 5-bromo-2-deoxyuridine (BrdU) 216
 incorporation during early S-phase for
 replication G-banding 224–5
 incorporation during late S-phase for
 replication G-banding 223–4
 5-bromo-4-chloro-3-indolyl phosphate
 (BCIP) 603
 Bruton's tyrosine kinase (Btk) 641
*Bst*XI site 447
Btk gene 641
 bulbar muscular atrophy 111
 bulked segregant analysis 785
 bulletin boards 813, 814–15, 885
 Burkitt's lymphoma 165

 c7dATP 558
 c7dGTP 558
 C-banding *see* constitutive
 heterochromatin banding
Caenorhabditis elegans 632, 688
 chromosome III 691, 692
 chromosome number 887
 cosmid clones 688
 DNA sequence 690
 expressed sequence tags 691
 genes 688
 genome 693
 map 688–9
 sequence 689–90
 size 888
 hybridization of individual YAC and
 cosmid clones 688
 mapping techniques 688–9
 physical map 690

- restriction enzyme-based fingerprinting 688, 689
- sequence-tagged site (STS) assays 688
- sequencing 513, 521
 - project 690–3
- targeted gene disruption 691
- transgenic reporter constructs 691
- YAC clones 688
- campomelic dysplasia 345
- cancer
 - chromosome aberrations 977–84
 - chromosome analysis 292
 - gene therapy 656–9
 - genetics and denaturing gradient gel electrophoresis 500
 - minimal residual disease detection 147
- capillary blotting, sequence transfer 608
- capillary electrophoresis 585–7
 - analysis time 586
 - band width 586
 - capillaries 586
 - electric field conditions 587
 - electro-osmotic flow 586
 - gel matrix 586–7
 - hydrophobic associations 587
 - Joule heating 586
 - polyacrylamide 587
 - polymer entanglements 587
 - sequencing gel 593–4
- carcinogenesis, rat model 296
- case–control tests 40
- CD4⁺ T cells 661
- cDNA
 - analysis in rice
 - large scale 776–8, 779–81, 782–3
 - strategy 777, 782
 - enrichment 443–4, 447–8, 458–62
 - filters for direct cosmid/YAC
 - hybridization 456–7
 - full-length 448
 - hybridization 443, 445
 - direct 447
 - identification in rice 778, 782–3
 - ligation to vector 443–5
 - minilibrary 443, 444
 - probes 217
 - rapid amplification 448, 462–6
 - selection 320
 - sequences 681
 - map-based cloning of target genes in rice 807–8
 - plant databases 811
 - sequencing in *Arabidopsis* nuclear genome 386–90
 - size fractionation 482–3
 - cDNA library construction 482–3
 - synthesis 480–2
 - cDNA library construction 480–2
 - from cytoplasmic RNA 452–4
 - transient expression screening 472
 - transfectant functional analysis 477, 491–2
 - walking 443, 444
- cDNA clones 103
 - enrichment factors 448
 - functional adhesion assays 491–2
 - large insert 448
 - rice 776, 778, 783
 - transient expression screening 471
- cDNA library 445
 - construction 472–6
 - cDNA size fractionation 482–3
 - cDNA synthesis 480–2
 - ligation of cDNA to vector 483–4
 - poly(A)⁺RNA preparation 479–80
 - protocol 478–86
 - rice 776–8
 - RNA isolation 478–9
 - vectors 472–6
 - viral genome replication 473
 - EBV-based episomal 475–6
 - functional adhesion assays on cloned cDNAs 491–2
 - rice
 - clone composition 782
 - genetic linkage map 786–801
 - redundancy analysis 782–3
 - tissue specificity 778
 - screening
 - cell-surface proteins 486–91
 - transient expression in mammalian cells 470–2
 - YACs 645
- cell culture 325–6
 - fusogens 326–7
- cell lines, unsynchronized cultures 277
- cell passaging 203–4
- cell-surface antigen identification 333
- cell-surface markers, fluorescent detection 248
- cell-surface molecules 472
- cell-surface protein screening by panning and rescue 476, 486–91
- cell-synchronizing agents 265–6
- CENSOR program 624
- Centre d'Etude du Polymorphisme Humain (CEPH) 53, 54, 55, 57, 90
 - genotyping errors 84
 - maps 46, 47, 48, 101
- cereal genome research 811–12
- CFTR* gene 652
- CFTR*deltaF508 mutation 9
- CG-clamping 502
- CGSC database 735
- charge-coupled device (CCD) 195, 606
- charge-coupled device (CCD) camera
 - Plate 8, 308–10
 - colour 310
 - dark current 309
 - photonic noise 309
 - pixel binning 309, 310
 - quantum efficiency 309
 - subarray sampling 309–10
 - types 309
- chemiluminescent substrates
 - alkaline phosphatase 604
 - enhanced for horseradish peroxidase 604–5
 - sequence labelling 604–6
- Chi sequence, *E. coli* 728
- CHIAS program 748
- chiasmata distribution 7, 8
- Chiasmotype Theory 7
- children, cytogenetic analysis 158, 159
- chimaera program 432, 433
- chimaeric clones 370
- chimaerism, long-range physical map
 - construction 432–3
- CHLC *see* Cooperative Human Linkage Centre (CHLC)
- chloramphenicol acetyl transferase (CAT) gene 674
- chloroplast DNA analysis 746
- chorionic somatotrophin, human genomic locus 626
- chorionic villus, unsynchronized cultures 277
- chorionic villus sample 174–7
 - harvesting 153–4, 177–9
 - prenatal diagnosis 153
 - transport medium 153
- chromatin
 - fibres
 - FISH probe mapping 222
 - release from interphase nuclei 215
 - interphase nuclei 216
 - release from interphase nuclei 225–6
 - released 217
- chromomycin A₃ dye 292, 293, 296, 297
- chromosomal DNA
 - denaturation 231–2
 - hybridization 232–3
- chromosomal *in situ* suppression (CISS) 218
- chromosome 6 microdissection 268
- chromosome 6q26–27 region 270
 - library 271
 - micro-FISH analysis of DNA Plate 7
- chromosome 21 cosmid contig Plate 8
- chromosome
 - aberrations 292
 - associated with cancer 977–84
 - abnormality
 - detection by chromosome painting Plate 6, 247–8
 - flow sorting Plate 5
 - malignant cells 159
 - analysis 147
 - bivariate 296, 297
 - lymphocyte preparation 152, 168–71
 - assignment for microcell-mediated chromosome transfer 339
 - bar codes 243
 - breakage 7
 - committees 881–5
 - cryopreservation 266
 - direct preparations 198–9
 - DOP-PCR amplification 253–5
 - Drosophila melanogaster* 669, 670, 671, 676, 677–8
 - Escherichia coli* 725–9
 - fixation 266
 - flow cytometry 190
 - flow sorting Plate 5, 147–8, 294, 299–300
 - flow-sorted
 - Alu-PCR amplification 252–3
 - DOP-PCR amplification 253–5
 - library 243–4

- construction 296, 298, 302–3
- markers
 - isozymes 753
 - RFLPs 753, 754
- metaphase 215–16
- number of common species 887
- preparation
 - for library construction 302–3
 - magnesium sulphate method 301–2
 - polyamine method for flow sorting 299–300
- probes *see* chromosome paints
- prometaphase 152
 - thymidine block synchronization 171–2
- ring 157, 159
- satellite 748
- segregation of somatic cell hybrids 328
- soup 170
- spreading 265
- walking 374
- see also* metaphase chromosomes; microdissection
- chromosome 17 345
- chromosome banding 266
 - karyotypic analysis 748–9
 - meiotic 753
 - microdissection 266
 - techniques 154–6
- see also* constitutive heterochromatin banding (C-banding); G-banding; quinacrine banding; R-banding
- chromosome painting 242–3, 295
 - applications 247–8
 - banding techniques 156
- chromosome abnormality detection
 - Plate 6
- chromosome libraries 243–4
- chromosome-specific probes 217
- commercial probes 246–7
- competitive *in situ* suppression
 - hybridization 242, 247, 255–7
- DOP-PCR amplification 253–5
- forward 295
- hybridized labelled probe detection 257–9
- interphase cytogenetics 248
- interspersed repetitive sequence-PCR 345
- limitations 248–9
- malignant myeloid disorders 249
- microdissection 249
- multicolour 246–7, 248
- nick translation 250–2
- probe concentrations 247
- probes Plate 4, Plate 5
- resources 244–7
- reverse 246, 249, 295
- chromosome paints 147
 - Alu-PCR 244–5
 - DOP-PCR 245
 - flow sorting of abnormal chromosomes 246
 - generation 298–9
 - IRS-PCR 244–5
 - microdissection and FISH 246
- chromosome pairing
 - cytoplasmic effects 751
 - discrimination between parental chromosomes 752–3
 - hybrid plants 749–51
 - marked chromosomes 752
 - mathematical models 753
 - meiotic chromosome banding 753
 - parental chromosomes 752–3
 - Ph1*-regulated 751–2
- chronic granulocytic leukaemia/chronic myeloid leukaemia (CGL/CML) 160, 162, 163, 164
- chronic lymphoblastic leukaemia 162
- chronic lymphocytic leukaemia (CLL) 161
 - symptoms 165
- chronic myeloblastic leukaemia t(9;22)
 - translocation 151
- chronic myeloid leukaemia 164
- chronic myelomonocytic leukaemia 160, 161
- Clarke–Carbon bank of *E. coli* 719, 724
- CLODSCORE 57–60
- clone, picking 326
- cloned DNA mapping 325
 - sequence 323
- cloning
 - map-based 802–3, 807–9
 - positional 369, 640
 - rings 326
- clotting factors 655
- cluster homology regions, yeast 708–9
- Cnx1 protein 780–1
- codA* gene 338
- codons 513
- colcemid 193
- Colibri database 735
- Collaborative Research (CRI) maps 101–2
- collagenase 191
- colorectal tumourigenesis 656
 - denaturing gradient gel electrophoresis 500
- colorimetric substrates, sequence labelling 603–4
- command line interface 808
- commercial suppliers 873–6
- comparative genomic hybridization (CGH) 193–5, 196
 - cameras 308
 - solid tumour cytogenetics 189, 206–9
- competitive *in situ* suppression
 - hybridization (CISSH) 242, 247, 255–7
- complementarity, local reverse 621, 623
- complex traits 28
 - analysis 38–40
 - genetic heterogeneity 33–5
 - Mendelian trait with covariates 28, 29, 30–3
 - no clear mode of inheritance 35–7
 - sampling problems 37–8
- comprehensive map 45
 - flowchart algorithm for construction 92
- computer
 - application programs 808
 - filing system 809
- mail lists 814
- network
 - connectivity 810
 - node 810
 - services 813–35
 - user interface 808–10
- computing
 - client–server 811
 - hardware 808
 - operating system 808
- consensus map for chromosome 1, MultiMap 94
- constitutive heterochromatin banding (C-banding) 155, 156, 159, 183–4
- karyotypic analysis 748–9
- contig 422
 - directed walking for long-range physical map construction 426
 - gap sequences 567
 - primers 567
 - shotgun sequencing 519
- contig assembly
 - Drosophila melanogaster* 678
 - microcloned DNA 273
- Cooperative Human Linkage Centre (CHLC) 47, 54, 57
- maps 101
- copia element of *Drosophila melanogaster* 673, 678
- COS cells 470, 472, 474
 - cDNA introduction 476
 - functional adhesion assays on cloned cDNAs 491–2
 - panning 476–7
 - transfectable 475
- cos sites 375
- cosmid 102–3, 217
 - clones
 - Caenorhabditis elegans* 688
 - host strain 376
 - colour determination 313
 - contigs 108
 - CsCl gradient 533, 535–6
 - purification 546–7
 - direct hybridization to cDNA filters 456–7
 - DNA 546–7
 - fingerprinting in *Drosophila melanogaster* 678
 - long-range mapping 370
 - QIAGEN plasmid kits 535
 - screening by hybridization 377–8
 - template amplification 533, 535–6
- cosmid libraries 370
 - clone handling 376
 - construction 397–401
 - Drosophila melanogaster* 677
 - host 375
 - insert DNA preparation 376
 - long-range mapping 375–6
 - rice physical map 805–6
 - vector 375
- costig program 430
- Cot-1 DNA 247
 - hybridization 443, 444
- CpG dinucleotide 444

- CRI-MAP 47, 49, 52, 53–4
 automatic map-building 81
 data set construction 62–3
 EUROGEM 85
 genotype file 51
 haplotype screening 84
 likelihood calculations 81
 locus file 63
 log₁₀ (likelihoods) 81
 MultiMap 91–2, 93, 94
 output 81
 parameter file 62, 64, 81, 84
 protocol 64–80
 reference maps 61–82
 use 63–4, 81
 CROP algorithm 83
 crossability 747
 crossing-over 6, 7
 double 7
 cryopreservation of blood and bone
 marrow samples 275–6
 cystic fibrosis
 chloride ion transporter (CFTR) protein
 650, 652
 expression vector 652–3
 denaturing gradient gel electrophoresis
 500
 gene therapy 652–3
 genetic markers 9
 genotypes 9
 linkage mapping 28
 phenotypes 9
 z₁ lod scores 14–16
 cytochalasin B 341
 cytogenetic analysis 147–8, 151–2
 applications 151–2
 approaches 167
 banding techniques 154–6
 cancer 159–61, 162, 163–4, 165, 166–7
 cell culture methodology 152–4
 constitutional abnormality detection
 156–9
 digital imaging 167–8
 microdissection 273
 prenatal diagnosis 153–4
 procedure 161, 163, 164, 165–6
 slide preparation 168–71
 whole blood sample processing 168–71
 cytogenetics
 nomenclature 161
 see also solid tumour cytogenetics
 cytokines 470, 472
 tumour cell expression 658
 cytomegalovirus enhancer 474
 cytosine deaminase 338
 cytotoxic genes 657–8

 D1S8 locus, MVR-PCR 135, 140–2
 DA/DAPI staining 155
dad1 gene 781
 DAPI
 digital camera imaging 310
 FISH 216
 fluorescence 314
 dark current 309
 data acquisition 812

 database 735, 736
 conceptual schema 812
 information sources 847
 organisms 844–6
 software 847
 subject indexes 844–7
 technology 811–12, 813
 data acquisition phase 812
 World Wide Web 847–61
 ddNTP *see* dideoxyribonucleoside
 triphosphate (ddNTP)
 dechimaerization inserts 432
 defective cell phenotype complementation
 470
 degenerate oligonucleotide primed PCR
 (DOP-PCR) 245–8, 280–1, 298
 amplification 270, 280–1
 microcloning of products 286–7
 chromosome painting 296
 flow-sorted chromosome amplification
 253–5
 microcloning of amplification products
 286–7
 microdissection 268–9, 270
 oligonucleotide primer 268–9
 denaturing gradient gel electrophoresis
 496–7
 applications 500
 band problems 508
 colorectal tumourigenesis 500
 constant 503
 DNA melting behaviour simulation
 498, 499
 DNA sequence analysis 496
 familial adenomatous polyposis 500,
 501
 GC-clamp 497–8, 499, 502
 genomic 502
 heteroduplex molecules 497, 501, 502
 homoduplex molecules 497
 molecular genetics 500, 502
 mutation detection 500, 503
 parallel 499
 perpendicular 498, 500
 plant genome analysis 786
 protocol 503–8
 resolving power 496–7
 SSCP analysis 503
 two-dimensional DNA typing 502–3
 variants of method 502–3
 deoxycytidine deaminase 333
 deoxycytidine kinase 333
 deoxyinosine (dITP) 558
 deoxyribonucleoside triphosphate (dNTP)
 561, 563, 564
 desynapsis 751
 DHFR gene 337–8
 diabetes mellitus, insulin-dependent
 (IDDM) 641
 genetic mapping 82
 dideoxy sequencing 514, 558, 560
 filamentous phages 562
 phagemids 563
 plasmid sequencing vectors 563
 vectors 562–3
 dideoxynucleotides, dye-labelled 610

 dideoxyribonucleoside triphosphate
 (ddNTP) 561, 563, 564
 differential replication banding
 early and late 155
 see also pre-banding
 digital microscopy
 colour reproduction 314
 FISH 30
 fluorescent chromosome band
 enhancement 313
 hardcopy output 314
 image data storage 314
 laser scanning 311–12
 multiple probe detection 312–14
 digoxigenin 218–19, 220, 228–30
 detection with FITC 235–6
 DNA labelling by nick translation
 228–30
 quality control of labelling 230–1
 sequence detection 617
 sequence labelling 602, 603
 digoxigenin-FITC Plate 9
 dihydrofolate reductase 337–8
 4th DIMENSION database 809–10
 2,4-dinitrophenyl (DNP) sequence
 labelling 602
 dinucleotide repeats 47, 108–9
 Généthon maps 101, 102
 loci 110–11
 typing 110–11, 119–21
 dioxetane substrates, chemiluminescent
 604
 diploid hybrids, plant genome analysis
 750
 direct transfer electrophoresis, sequence
 transfer 608
 disaggregation
 enzymatic 191, 197–8
 mechanical 190–1, 196–7
 disease gene
 LINKMAP 83
 mapping 81–2
 disease susceptibility allele 37
 disease-resistance gene tracking 779
 dispersed repeat element pre-association
 114–15
 distamycin A (DA) 155
 DNA
 extraction, from P1 clones 414–16
 fragment denaturing gradient 496
 gyrase binding in *E. coli* 719
 high molecular weight
 liquid preparation 380
 preparation in agarose 378–9
 isolation from agarose plugs 226, 227
 ligase 106–7
 melting behaviour simulation 498, 499
 mismatch repair genes 500
 packaging 370
 polymerase 500, 566
 reduced fingerprint 134
 single-copy content of genome 888–9
 single-stranded template 569–70
 universal amplification 269
 see also dideoxy sequencing; insert DNA;
 sequence; sequencing

- DNA damage 265
 acid-induced 266
 mitotic spindle inhibitors 266
 sources 265–6
 synchronizing agents 265–6
- DNA fingerprinting 3, 113, 128, 129, 130
 data interpretation 132
 linkage analysis 131
 multilocus 131
 practice 132
 probes 131
 protocol 136–7
 statistics 132
- DNA markers
 mapping 325
 mouse genome mapping 635
 rice linkage analysis 803, 804
 rice physical map 806, 807
- DNA polymorphisms 99
 A residues 110
 applications 99
 CEPH maps 101
 CHLC maps 101
 classes 99
 cosmids 102–3
 defined clone 102–3
 dispersed repeat element pre-
 association 114–15
 EUROGEN maps 101
 finding 100–2
 Génethon maps 101
 identifying new 102–12
 informativeness 99–100
 large collections 107–12
 map placement 100, 100–1
 multilocus methodology 113
 screening for 110
 sequence analysis 104–6
 sequenced genes 103
 Southern blot hybridization 102, 103–4
 SSCP analysis 103, 104, 105, 115–17
 tandem repeat variability 108–9
 YACs 102–3
- DNA probes
 denaturation 232–3
 detection of hybridized 234–7, 257–9
 mapping/ordering by FISH Plate 3
 microdissected region-specific 282–6
 plant genome analysis 754
 post-hybridization washes 233–4
 prehybridization 232–3
 rice genetic linkage map 786–801
- DNA typing 128
 allele frequencies 130
 anonymous coding 131
 cell line identity 128
 confidentiality 131
 DNA sample quality control 128
 family relationship verification 128
 hypervariable minisatellite loci 132
 locus-specific minisatellite probes
 132–4
 paternity analysis 128, 130
 statistical evaluation 129–30
 systems 128–9
- DNASTAR program 733
- dNTP *see* deoxyribonucleoside
 triphosphate (dNTP)
- DOMAINER program 771
- dominant selectable markers 334–5
 eukaryotic expression vectors 335–6
 negative selection 338
 positive selection 335–8
 positive–negative bidirectional selection
 338
 promoter 334
 selection schemes 335–8
 transfer into mammalian cells 334
- DOP-PCR amplification *see* degenerate
 oligonucleotide primed PCR
 (DOP-PCR)
- double minutes 189
 alveolar rhabdomyosarcoma 195
- double-cos vectors 375
- doubled haploid lines 786
- Drosophila* Genome Centre mapping
 project 674, 679–82, 683
in situ hybridization mapping 680
- Drosophila melanogaster* 632, 668
 Bridge's map 670, 671–2
 cDNA sequences 681
 chromosome number 887
 clone ordering by *in situ* hybridization
 674, 675
 contig assembly 678
 copia elements 673, 678
 cosmid clone availability 679
 cosmid fingerprinting 678
 cosmid library construction 677
 cytogenetic mapping 669, 671–2
 cytogenetics 669, 670, 671–2
 Duncan map 676
 euchromatin 669, 671
 European Consortium cosmid map 674,
 676–9, 683
 evolutionary relationships 672
 FlyBase database 681, 682
 foldback DNA 673
 genetic mapping 668–9
 genome sequencing 682
 genome size 888
 genome structure 672–3
 Hartl map 676
 heterochromatin 669
 histone genes 673
 large-scale sequencing 681
 long-terminal inverted repeat elements
 673
 mapping projects 674–83
 mitotic chromosomes 669, 670
 model system 668–9, 670, 671–4
 molecular genetics 672–4
 mutations 668–9
 P1 clone availability 681–2
 P1 clone library 676
 P1 library 680
 P element 673–4, 680–1
 insertions 679
 P element-mediated germline
 transformation 674
 polytene chromosomes 669, 670, 671–2,
 676, 677–8
- ribosomal RNA genes 672–3
 satellite DNA 672
 sequence tagged sites (STSs) 678–9,
 680–1
 transposable elements 673
 transposon tagging 674
 YAC maps 674, 675–6
- Drosophila pseudoobscura* 683
Drosophila viridis 682–3
 dwarf gene, cereals 811–12
 dystrophin gene 10
- E. coli* F factor 371
 e-mail 812, 814
 EBNA1 replication protein 475
 ebnavirus 473
 ECO2DBASE 736
 EcoCyc database 735
 EcoGene database 735
 EcoMap database 735
 EcoSeq database 735
 electro-osmotic flow 586
 electrophoresis 578–9
 band width 578, 579
 band-broadening 578
 band-spacing 578, 580
 capillary gel 585–7, 593
 diffusion 578, 579, 580
 direct blotting gel 579, 590–2
 direct transfer 584–5
 gel for automated sequencing 592–3
 Joule heating 578, 579, 580
 limiting factors 580
 molecular orientation 578, 580
 pulsed electric fields 579
 resolution 578
 sequence transfer 608
 slab gel 579–85
 standard sequencing gel 588–90
- Encyclopedia of the Mouse Genome
 639–40
- enhancer trap, *Arabidopsis* nuclear genome
 769
- enterobacterial repetitive intergenic
 consensus (ERIC) sequence 728
- enucleation 357
 from plastic bullets 351–3
 Percoll gradient 353–4
- epifluorescence microscopy 306–11
 arc lamp 306–7
 CCD camera 308–10
 digital cameras 307
 digital image 307
 electronic cameras 307–8
 image digitization 311
 Köhler illumination optics 306, 307
 linear filter 313
 multiple fluorochrome imaging
 310–11
 multiple probe detection 313
 objective lens 307
 ratio labelling 313
 silicon intensified camera 308
 video cameras 307, 308
- Epstein–Barr virus 152, 473, 660
 transformation of blood cells 871

- vector for cDNA library construction 473, 474
- Escherichia coli* 632, 717
- Arabidopsis* genes cloned 767
- base composition 726
- Chi sequence 728
- chromosome 720, 725–9
 - number 887
 - replication 726
- circular chromosome 718
- Clarke–Carbon bank 719, 721
- codon usage 726
- Colibri 735
- complementation 724
- complementing fragment direct identification 724
- conjugation 723–4
 - followed by complementation 724
- conserved sequences 738
- contigs 725
 - editing 733
- cotransduction frequency 724
- CTAG sequence 726
- databases 735–6
- DNA 719, 720, 723
 - gyrase binding 719
 - replication 726
 - sequence 725–6
- DNA polymerase I 561
 - binding 719
- EcoCyc 735
- EcoGene 735
- Ecomap5 721
- EcoMap 735
- Ecoseq 735
- G+C content 725, 734
- gene
 - analysis 733–4
 - annotation 733–4
 - arrangement 725, 726–7
 - order conservation 727
 - products 726
 - rearrangement 727
- Gene Database 736
- Gene-protein Database 736
- genetic mapping 724–5
- Genetic Stock Centre (CGSC) 735
- genome 718–19, 720
 - nomenclature 717
 - segmentation 734
 - sequence 710
 - sequencing projects 729–34
 - size 888
- gray holes 734
- Harvard University genome sequencing 730–1
- Hfr strains 723
- I-SceI digestion 734
- insertion sequences 728
- interspersed repeats 727–8
- inversions 727
- K-12 718, 737–8
- Kobe University genome sequencing 730
- Kohara clone library 719, 721
- Kohara map 724
- lambda clones 719
- lambdoid phages 728
- life cycle 717
- linkage map 722
- Moco (molybdeum cofactor) mutant 781
- model system 717–18
- open reading frames 725, 733
- physical map 719, 721–2, 722
 - methods dependent on 724–5
- plasmids 719
- potential gene identification in
 - provisional sequence 733
- promoter location 733
- random clones for sequencing 731
- REP sequences 719
- repeated sequences 727–9
- replication 719
- replicons 719
- reverse genetics 724–5
- Rhs elements 728
- ribosomal protein 726
- ribosomal RNA 726
- sequence similarity 725
- sequencing process 732
- short multicopy palindromic repeats 728–9
- shotgun preparation 731
- SWISS-PROT 736
- termination 719
- terminus region 726
- Tokyo University genome sequencing 730
- transcriptional unit orientation 727
- transduction 723, 724
- transposable element Tn10 724
- transposon transmission 728
- Wisconsin University group genome sequencing 730, 731–4
- essential thrombocythaemia 166
- ethernet 810
- ethidium bromide 296
- ethotrexate 337
- euchromatin, *Drosophila melanogaster* 669, 671
- EUCIB mouse-backcross map 425–6
- eukaryotic expression vectors 335–6
- eukaryotic genes, functional analysis 631
- EUROFAN 711
- EUROGEM maps 44, 45, 46, 47, 101
- European Collaborative Interspecific Backcross (EUCIB) programme 638, 640
- European Consortium cosmid map 674, 676–9, 683
- European Gene Mapping Project (EUROGEM) 85
- European Scientists Sequencing *Arabidopsis* (ESSA) project 770, 771, 774–5, 778
- Ewing’s sarcoma, media for cytogenetics 196
- exon amplification 443–4
 - artefacts 446
 - efficiency 445–6
 - specificity 445–6
- exon DNA cloning 454–5
- exon trapping 320, 442–3
 - chimaeric exons 446
 - protocol modification 447
 - pSPL1 445–7, 449–56
- exonuclease III 523, 524, 536, 558
- expectation maximization (EM) algorithm 49, 92
- expressed sequence tags (ESTs) 319, 514
 - Arabidopsis*
 - genetic maps 765
 - nuclear genome 769–73
 - Caenorhabditis elegans* 691
 - mapping 320
 - rice 776
 - cDNA clones 783
 - marker for genetic linkage map 802
- extracellular proteins, transient expression screening 477
- F2 progeny 786
- factor IX 655
- familial adenomatous polyposis (FAP)
 - APC gene 503
 - mutations 500, 501
 - de novo* mutations 10
 - linkage mapping 28
 - mutation penetrance 10
 - phenotype variation 9–10
- family
 - genetic map data 48–50
 - inbred 38
 - loops 38
 - structure 23
- Fanconi’s anaemia 476
- FASTMAP program 49
- fibre distributed data interface (FDDI) 810
- file transfer (ftp) 815–33
- FITC *see* fluorescein isothiocyanate (FITC)
- FKHR gene Plate 2
- flow cytometer, dual-laser 296
- flow cytometry 292
 - chromosomes 190
 - type differentiation 294
- flow karyotype 292
 - bivariate 293
 - human fibroblast chromosomes 294
 - variations 292–3
- flow karyotyping of chromosomes 298
- flow sorting 108
 - abnormal chromosomes Plate 5
 - chromosome preparation
 - magnesium sulphate method 301
 - polyamine method 299–300
 - with PCR 294
 - see also* fluorescence-activated flow sorting
- flow-sorted chromosomes
 - Alu-PCR amplification 252–3
 - DOP-PCR amplification 253–5
- fluorescein, sequence labelling 602
- fluorescein isothiocyanate (FITC) 215
 - biotin detection 235
 - digital camera imaging 310
 - digoxigenin detection 235–6
 - PCR product signal development 296

- fluorescence
banding techniques 313
plus Giemsa method 216
- fluorescence *in situ* hybridization (FISH)
147, 215
alpha satellite probes 217–18
Alu-PCR 227–8
applications 215
banding 216
biotin 218–19, 220
labelling 228–31
biotin-labelled probes 243
BrdU incorporation 223–5
chromatin release 222, 225–6
chromosomal DNA
denaturation 231–2
hybridization 232–3
chromosomal *in situ* suppression (CISS)
218
chromosome microdissection
combination 242
digital microscopy 30
digoxigenin 218–19, 220
labelling 228–31
digoxigenin-labelled probes 243
DNA
isolation from agarose plugs 226, 227
labelling 228–30
probe mapping/ordering Plate 3
fluorescence plus Giemsa method 216
fluorochromes 877–8
hybridized probe detection 234–7
interphase 195
analysis Plate 2, 189
cells 248
nuclei 216–17
long-range physical map positional
information 424
metaphase chromosomes 215–16
microdissected region-specific probes
282–6
microdissection 246, 282–6
monosomy 7 in leukaemia 250
multicolour techniques Plate 9, 248
multiple probe detection 312–14
nick translation 219, 228–30, 250–2
pepsin pretreatment 219–20
plant genome analysis 754, 755
post-hybridization washes 233–4
pre-banding 222
probe 217–18
detection 220
DNA denaturation/pre-hybridization
232–3
labelling 218–19
mapping 220–2
signal analysis 220–2
suppression 218
proteinase K pretreatment 219–20
R-banding 216
released chromatin 217
replication G-banding 223–4
replication R-banding 224–5
slide preparation 215–17
solid tumour cytogenetics 189, 193–5
somatic cell hybrid characterization 344
- TRAP gene mapping 221
whole chromosome painting probes
Plate 4, Plate 5
YACs 218
see also chromosome painting
- fluorescence-activated flow sorting 292–6
analysis 242
chromosome
library construction 298
paint generation 298–9
preparation 298, 299–302
fluorochromes 292, 296
instrumentation 296–7
see also flow sorting
- fluorogenic substrates, sequence labelling
606
- fluorochrome-conjugated nucleotides 218,
219
- fluorochromes 147, 215, 877–8
fluorescence-activated flow sorting 292,
296
imaging with digital camera 310–11
laser scanning microscopy 312
multiple 313
probe labelling 219
5-fluorocytosine/cytosine deaminase 338
FlyBase database 681, 682
folate deficiency 152
foldback DNA, *Drosophila melanogaster*
673
- forensic practice
application of DNA typing 128
minisatellite loci band-shift effect 133
- form filling 808
- fragile X syndrome 10
cytogenetic analysis 159
detection 152–3, 172–3
trinucleotide repeats 111
- framework map 45
flowchart algorithm for construction 91
- fungal infection 326
- fusogens 324, 326–7
- G418 sulphate/neomycin
phosphotransferase II 335–6
- G418-resistance gene 675
- G-11 banding 155
- G-banding 152, 154–6
breast carcinoma 192
FISH 223–4
karyotypic analysis of somatic
chromosomes 748–9
microdissection 266
protocol 179–80
restriction endonuclease 155
- Ga14 promoter 674
- GAMA medium 333
- ganciclovir 658
- ganciclovir/herpes simplex virus
thymidine kinase 338
- GC-clamp 497–8, 499
- gel digest fingerprinting 422, 423
- gel electrophoresis, sequencing 572
- gene
expression regulation 631
function 631
- fusions 147
mapping with microcloned DNA 273
tagging for map-based cloning 802–3
transfer
accuracy 661–2
adenoviral vectors 662
- gene isolation from genomic DNA 442–5
cDNA enrichment 443–4, 447–8, 458–62
coding sequence isolation 444–5
CpG islands 444
direct cDNA hybridization 447
direct cosmid/YAC hybridization to
cDNA filters 456–7
exon amplification 443–4
exon trapping 442–3
by pSPL1 445–7, 449–56
5' RACE 464–6
rapid amplification of 3' ends 462–4
- gene therapy 650
ADA deficiency 662, 663
AIDS 652, 660–1
cancer 656–9
candidate disease 650–2
corrective 657
cystic fibrosis 652–3
cytotoxic 657–8
delivery 652–3
systems 661–3
vehicle 651
- ethics 650
- germline 650
- heritable potential 650
- HIV infection 660–1
- immunotherapy 657, 658–9
- infectious disease 659–61
- monogenic disorders 650–1, 652–6
- multifactorial genetic disorders 656–9
- physiological defect correction 651, 652
- regulation of expression 651
- severe combined immune deficiency
(SCID) 653–5
- somatic 650
- target cells 651–2
- thalassaemia 654, 655
- vector delivery 661
- vectors 652–3, 653–4
- viral infection 659–61
- GENE/COMBIS electronic journal 805
- Génethon human genome map 101, 102,
425
markers 101
- genetic character segregation
meiotic 6–7
non-independent 6
- genetic disorders
chromosomal changes 160
multifactorial 656–9
translocation 147
- genetic distance measuring 6–7
- genetic mapping
Drosophila melanogaster 668–9
heuristics 90
- genetic maps 44–7
disease mapping 81–2
family data 48–50
high-density linkage 3–4

- integration with physical 85
- interference 50
- likelihood 48
- locus distance 44
- lod scores 49
- marker position 49
- MultiMap 90–5
- multiple two-point analyses 49–50
- multipoint 7, 52
- order inference 47–8
- polymorphisms 47
- probability computing 48–50
- protocols 50–84
- recombination information 47–8
- sex differences 44
- shared resources 46
- types 45
- genetic markers 8–9
- genetic polymorphisms 3, 99
- genetic recombination 7
- genetic variation 3
- genome
 - comparison 631
 - definition 746
 - mapping
 - Diptera 682–3
 - random clone anchoring 433–8
 - mismatch scanning 113
 - scanning of *E. coli* K-12 737
 - sequence of *Caenorhabditis elegans* 689–90
 - size 888–9
 - see also* human genome map
- Genome Data Base (GDB) 57, 812
 - allele frequency 136
- genome resource centres 846–7
 - World Wide Web 847–61
- genome-restructuring genes 755
- genomic DNA
 - cloning 447
 - exon trapping by pSPL1 449–56
 - labelling by nick translation 206–9
 - preparation 372
 - production of clonable 383–4
- genomic hybridization analysis
 - alveolar rhabdomyosarcoma Plate 1
 - see also* comparative genomic hybridization (CGH)
- genomic *in situ* hybridization, plant
 - genome analysis 754, 755
- genotoxicity assays 295
- genotype 9–10
- genotyping errors 93
- GenProTec database 736
- GenQuest server 624
- germ cell tumours, media for cytogenetics 196
- germline
 - mutation rate 133
 - P element-mediated transformation 674
- gestation, cytogenetic analysis 156–7, 159
- Giemsa banding *see* G-banding
- gliadin profiles 753
- global repeats 621, 624
 - inverted 621–3
- β -globin chains 654, 655
- globin gene family transcriptional control 655
- Glrh* gene 640
- glycerol kinase deficiency gene 445
- Gopher 816, 833
- gpt* gene 336, 338
- granulocyte-macrophage colony-stimulating factor (GM-CSF) 472
- graphical user interfaces (GUIs) 808, 809, 811, 812
- grasses
 - classification 747
 - synteny 811–12
- gray holes, *E. coli* 734
- growth hormone, human genomic locus 626
- haematological malignancy 160, 162
 - genes in translocations/inversions 162
 - symptoms 164–6
- haemophilia, denaturing gradient gel electrophoresis 500
- haemophilia A 22
- haemophilia B 655
- Haemophilus influenzae*
 - genome sequence 736, 737
 - genome size 888
- hairy-cell leukaemia (HCL) 165
- haplome 746
- haplotypes
 - analysis 21–2
 - construction 24
- hapten, probe labelling 218
- HAT medium preparation 346–7
- HAT selection 332–3
- head and neck tumour, media for
 - cytogenetics 196
- hepatitis B 660
- hepatocellular carcinoma 660
- hereditary non-polyposis colon cancer (HNPCC) 10
- herpes simplex virus thymidine kinase
 - gene (HSV1/k) 338, 658, 659, 660–1
- heterochromatin 669, 671
- heteroduplex formation 112
- heterogeneity 33–5
- heterokaryons 324
- heterozygosity 99–100
- heuristics
 - genetic mapping 90
 - probe ordering 429–30
- Hga*1 107
- Hinfl* fragment 104, 105
- hisD* gene 337
- histidinol/histidinol dehydrogenase 337
- histone genes 673
- HIV infection, gene therapy 652, 660–1
- HLA-DQ-locus 128
- Hodgkin's disease 165
- Hoechst 33258 dye 292, 293, 296, 297
- homeobox genes 780
- homogenously staining regions 189
- homologous recombination 334
- Hordeum vulgare*, banding analysis 749
- horseradish peroxidase
 - enhanced chemiluminescent substrates 604–5
 - sequence labelling 603
- host, computer network node 810
- hph* gene 336
- HPRT* gene 338
- hsp70* promoter 675
- HUGO chromosome committees 881–5
- human chromosome
 - endogenous selection genes 329–31
 - size 887
- human disease genes 879, 880–922
 - positional cloning 325
 - yeast gene similarities 710
- human gene mapping 8–16
 - de novo* mutations 10
 - family data 9
 - family studies 8–10
 - genetic markers 8–9
 - genotypes 9–10
 - linkage analysis 10–11
 - linkage maps 8
 - penetrance 10
 - phenocopies 10
 - phenotypes 9–10
- human genome 513–16
 - morbid anatomy 923–42
- Human Genome Data Base (GDB) 90
- human genome map 45
 - Généthon 101, 102, 425
- Human Genome Mapping Project 51
- Human Genome Project, model organisms 634
- HUMGHCSA genomic region 626–7
- Huntington's disease 10
 - gene 442
 - linkage mapping 28
 - trinucleotide repeats 111
- hybrid phenotype mapping 323
- hybridization
 - locus-specific probes 103
 - multiple-copy probe 422–3, 424, 425
 - phenotypic changes 338
 - sequence detection 600, 611–12
 - sequencing by (SBH) 568–9, 573
 - single-copy probe 422, 423–4, 425
 - see also* comparative genomic hybridization (CGH)
- hygromycin B/hygromycin B kinase 336
- hyperekplexia 641
- hypertext markup language (HPML) 834
- hypoxanthine phosphoribosyltransferase 332–3
- icons 809, 810
- identity-by-descent 36
- identity-by-state 36
- IGD/X-PED system 53, 54–6, 57, 58–60
 - chromosome screen 57
 - data management 57
 - disease gene pinpointing 83
- image analysis 306
- immunomodulatory gene
 - delivery 657, 658
 - expression 657
- immunotherapy, gene therapy 657, 658–9

- Imperial Cancer Research Fund, WWW server 817
- in situ* hybridization
plant genome analysis 754
see also fluorescence *in situ* hybridization; genomic *in situ* hybridization
- inbreeding loops 38
- inclusive map 45, 92
- infectious disease, gene therapy 659–61
- infertility 159
- inner product mapping 425–6
- insert DNA
phosphatase treatment of partial digests 404
preparation for cosmid libraries 376
- Integrated Genomic Database (IGD) 47
project 52–3
public data 56–7
- Integrated Services Digital Network (ISDN) 810
- integration of genes into chromosomes 334
- integrins, host cell 471–2
- intensified silicon intensified target camera 308
- interference 7, 8, 50
- intergenic repeat unit (IRU) sequences 728
- intergenomic affinity, plant genome analysis 750
- internal repeat recognition 624–5
- Internet 810
commercial service provision 810
resources 805
- interphase cytogenetics 248
- interphase nuclei 216–17
chromatin release 225–6
detection rate 248
ordering of FISH probes 221–2
structural chromosomal abnormalities 248
- interspecific hybrids 324
- interspersed repeat sequence PCR (IRS-PCR) 242, 244–5, 344–5
EUCIB mouse-backcross map 425
high-resolution genetic mapping 638
markers 644
seed contig extending 644–5
segregation analysis in mouse 635
YAC clones 374–5
- interspersed repeat sequences, physical maps 644–5
- intracellular proteins
screening by *in situ* labelling 477
transient expression screening 471
- intron-exon structure 103
- introns, yeast 701–2
- irradiation and fusion gene transfer (IFGT) 324, 325, 342–4
fusion 343
selection 342–3
- irradiation fusion hybrids 319
- IS elements, *E. coli* 728
- isozymes 753
- JANET 810, 811
- Japanese Rice Genome Research Program 776
- JOINMAP program 764
- journals 885
electronic 805
- kanamycin resistance 375
- karyoplasts 341–2
- karyotypic analysis
chromosome banding 748–9
computer-aided 748
conventionally stained
somatic/pachytene chromosomes 747–9
Giemsa-banded somatic chromosomes 748–9
mitotic 748
pachytene 748
satellited (SAT) chromosomes 748
- karyotyping 167–8
- kinetochore staining 155
- Klenow fragment 558, 560, 566
enzymatic DNA sequencing 563–4, 565
sequence analysis of PCR products 571
- knock-out tables 943, 944–56
- Kohara clone library of *E. coli* 719, 721, 724
- Kohara map of *E. coli* 724
- Köhler illumination optics 306, 307
- label multiplexing 606
- lacZ* gene 674
- lacZ* promoter 562
- laser scanning microscopy 311–12
confocal 312
- laser scanning technology 306
- leukaemia
acute 164
monosomy 7 250
see also acute and chronic leukaemias
- liability classes 30
breast cancer 31, 32
- ligation, template-directed 568
- light* gene 669
- light-signalling pathway 780
- likelihood 48
- likelihood function, CRI-MAP 61–2
- linkage 6
heterogeneity 33, 34–5
- linkage analysis
affected individuals 32–3
age 32
breast cancer 30–3
efficiency 39
human 10–11
IGD/X-PED system 45–6, 54–6
lod scores 11–12
phenotypic heterogeneity 34
polymorphic markers 39
recombination fraction 31, 33
risk of disease 33
unaffected individuals 32–3
- linkage map
consensus 93
construction 90
human 8
- linkage markers
not segregating in Mendelian pattern 23–4
segregation with disease 24
- LINKAGE program 11, 12, 17, 22
alleles 39
disease mapping 82, 83
files 50
input files 57–61
liability classes 30
probands 38
- linkage studies
haplotype analysis 21–2
human 17, 21–2
markers 17, 21
mode of inheritance 37, 38
number of families 37–8
- linker-adaptor PCR (LA-PCR) 244, 269
- LINKMAP disease mapping 82, 83
- LIPED program 11, 17
- Lipofectin 192, 202–3
- LISP interpreter 93, 94
- local area networks 810, 811
- locus control region (LCR) 655
- locus distance 24, 44
- lod scores 11–12
breast cancer 34, 35
calculation 12–16
CRI-MAP 62
data collection 17, 18, 19–20, 21–2
LINKAGE input file creation 60–1
mapping 49
phase-known vs. phase-unknown
linkage data 17
tables 17, 18–21
- long interspersed repeat elements 345
- long-range physical map construction 369, 422
chimaeric clone detection 430, 432–3
cloning system 369–71
competition of probe to remove repetitive sequences 416–17
computational requirements 426–7
cosmid libraries 375–6
construction 397–401
cosmids 370
data 422–4
entry 426
visualization 427
error checking 426
experimental strategy 426
feedback 426
fitting clones to probe order 430
genetic and physical information integration 425–6
genomic DNA preparation 372
high molecular weight DNA preparation 378–80
human chromosome 422
hybridization 377–8, 417–18
matrix 422
inner product mapping 424–5
least distant neighbour 429, 430
library resources 369
map forking 429–30
most distant neighbour 429, 430

- multiple-copy probes 422–3, 425
- neighbourhood rules 429, 430
- P1 clones 369, 370–1
 - DNA extraction 414–16
- P1 library construction 376–7, 401–14
- positional information 424, 427
- probe
 - contig 427
 - ordering 427–9
- random noise detection 430, 432–3
- screening by hybridization 377–8
- single-copy probe 422, 425
 - data 427–30
- somatic cell hybrids 369
- washing 417–18
- YAC 369, 370
 - agarose block preparation 373, 391–3
 - DNA partial restriction digest mapping 394–5
 - filter lifts 389–91
 - generation of end-specific probes from clones 395–7
 - library construction 372, 381–8
 - size fractionation by PFGE 393–4
- long-terminal inverted repeat elements,
Drosophila melanogaster 673
- Lophopyrum elongatum*
 - C-banding 749
 - chromosome pairing 751, 752
- lung tumours, media for cytogenetics 196
- lymphoblastoid cell lines, flow karyotypes 294
- lymphocyte
 - mitogens 152
 - separation of peripheral blood samples 274
 - sterile separations 870
 - transformation 870
- lymphokine-activated killer (LAK) cells 658
- lymphoma 160
 - chromosome abnormalities 981
- lymphoproliferative disorders
 - chromosome changes 982
 - chronic 160
- M13 DNA
 - cloning vector 561–2
 - detergent extraction 542–3
 - magnetic bead purification 543–4
- M13 template 533, 534–5, 558
 - detergent extraction method 533, 534, 542–3
 - magnetic bead purification 533, 534–5, 543–4
 - silica gel membrane purification 533, 535, 545, 545
 - standard PEG-phenol method for DNA recovery 533, 534, 541
- magnesium sulphate, chromosome preparation for flow sorting 298, 301–2
- mail lists 814
- malignancy, cytogenetic analysis 159
- malignant cell
 - chromosomal abnormalities 159
 - transformation 656
- malignant hyperthermia susceptibility 22
- map distance 8
- map placement, DNA polymorphisms 100–1
- map-based cloning 802–3
 - target genes 807–9
- MAPMAKER
 - mouse genome mapping 637
 - rice genetic linkage map 803
 - rice genome mapping 785
- mapping function 7–8
- marker alleles 39–40
- marriage loops 38
- Massachusetts Institute of Technology (MIT) microsatellite map 637, 640
- materials 863–7
- matrix-assisted laser desorption 573
- Maxam–Gilbert sequencing method 559–61
- Maximum Likelihood Estimate 48, 49
- Maximum Likelihood Order 48
- MBx database 640, 643
- media 863–7
- meiosis
 - gene exchange 6
 - phase-known 17
 - phase-unknown 17
- melting domain 498, 499
 - genomic denaturing gradient gel electrophoresis 502
- melmap of DNA fragment 498, 499
- Mendelian trait with covariates 28, 29, 30–3
- Menkes disease gene 442
- menus 808
- mesenchymal tumours, media for cytogenetics 196
- met oncogene 333
- metaphase chromosomes
 - localization of FISH probes 220
 - ordering of FISH probes 221
- methotrexate/dihydrofolate reductase 337–8
- micro-FISH analysis 270, 273
 - chromosome region 6q26–27 DNA Plate 7
- microcell hybrids 324, 325
 - applications 323
- microcell-mediated chromosome transfer (MMCT) 324–5, 339–42
 - colcemid conditions for micronucleation 341
 - filtration 341–2
 - fusion 342
 - human monochromosomal hybrids 334, 335
 - karyoplasts 341–2
 - microcell isolation/enucleation 341
 - micronucleation 340–1
 - timetable 340
- microcells
 - filtration 354–5
 - fusion to whole recipient cells 355–7
- microclones
 - characterization 271–3
 - chromosomal regional specificity 272–3
 - frequency determination of repetitive and unique 272
 - human origin confirmation 272–3
 - insert
 - colony PCR 287–8
 - isolation by colony PCR 287–8
 - size 271
 - isolation of potential candidate genes in disease 273
 - library construction 270–3
 - redundancy level 272
 - Southern blot analysis 272
- microcloning 273
 - chromosome harvesting 277–8
 - DOP-PCR amplification products 286–7
 - microdissection 268–9
- microdeletion syndromes 242
- microdeletions 158
- microdissection 148, 265
 - amplification reaction 270
 - applications 265, 273
 - cell culture techniques 266
 - chromosome
 - harvesting 277–8
 - preparation 265–6
 - chromosome 6 268
 - colony PCR 287–8
 - contamination 269–70
 - cryopreservation 275–6
 - cytogenetic analysis 273
 - DOP-PCR amplification reaction 280–1
 - equipment 267
 - FISH 246, 282–6
 - libraries 108
 - lymphocyte separation of peripheral blood samples 274
 - micro-FISH analysis 270
 - microcloning 268–9, 286–7
 - microneedle preparation 267, 278–9
 - microscopic 265, 266–8
 - plant genome analysis 755
 - region-specific libraries 273
 - technique 279–80
 - translocation breakpoints 273
 - universal DNA amplification 269
 - unsynchronized cultures 276–7
- micronucleation 356–7
 - optimization 350
- microsatellite map
 - MIT 637, 640
 - mouse genome mapping 637, 640
 - polygenic loci 640–1
- microsatellites
 - repeat polymorphisms 46–7
 - see also short tandem repeats
- Miller–Dieker syndrome 242
- minisatellite variant repeat PCR (MVR-PCR) 129, 130, 134
 - data interpretation 135
 - protocol 140–2
- minisatellites
 - allele drop-out 134
 - allele frequencies 133–4
 - band-shift effect 133
 - duplication 110

- hybridization screening 112
hypervariable loci 132
locus 100
 properties 132
 typing 112, 128–9, 133–4
locus-specific probes 132–4
 protocol 138–40
match criteria 133
PCR analysis 129, 134–5
probes for screening 112
single locus typing 128–9
variability 112
minisequencing, solid-phase 107
mitochondrial DNA typing 128
mitogens 166, 167
mitotic karyotyping 748
mitotic spindle inhibitors 266
MLH1 gene 500
modem 810
mogA 781
monogenic disorders 650–1
 complex 655–6
mononucleotide repeats 110
monosomy 7 in leukaemia 250
mouse gene knock-out tables 943
 double knockouts 955–6
 targeted mutations 944–55
mouse genetic map
 comparative with human genome 641
 polygenic loci 640–1
 using 640–1
Mouse Genome Database 639–40
mouse genome mapping 634
 accessing information 639–40
 back-cross analysis 636–7
 candidate gene confirmation by
 positional cloning 645
 current status of map 637–8
 DNA markers 635
 high-resolution map 638–9
 interspecific/intersubspecific genetic
 cross 635–6, 637
 methodologies 635–7
 microsatellite map 637, 640
nomenclature 640
physical 641–5
 clone resources 641–2
 contig closure 644
 databases 643
 high-resolution genetic maps 643–4
 interspersed repeat sequences 644–5
 uses 645
publications 640
recombinant inbred strains 637
strain distribution pattern 637
strategies 634–41
mouse mutation 634
 candidate gene identification 640
 mapping as prelude to positional
 cloning 640
MSH2 gene 500
multicomponent complexes, transient
 expression screening 471–2
multidrug resistant protein (MDR-1) 659
multifluorochrome techniques 306
MultiMap 90–5
 consensus map for chromosome 1 94
 CRI-MAP 91–2, 93, 94
 documentation 94
 LISP interpreter 93, 94
 locus markers 91
 mailing list 95
 software 94
 troubleshooting 95
 uses 92–3
multiple myeloma 166
multiplex sequencing 524–6
 chemiluminescence reaction 525
 hapten labels 526
 probe labelling 524–5
 sequence detection 611–12
 single vector
 with multiplex probe labelling 525–6
 with multiplex tagged primers 525
 streptavidin bridge 525
 tagged vectors 525, 526
multiplex tagged vectors 525
Mus castaneus 635, 637
Mus spretus 635, 636, 638
mutagenesis, rat model 296
mutation detection 496
 denaturing gradient gel electrophoresis
 500, 503–8, 503
 mutations, *de novo* 10
MutHLS mismatch repair proteins 113
myb oncogenes 780
MycDB 835
Mycobacterium database 835
mycophenolic acid/xanthine
 phosphoribosyltransferase 336
mycoplasma 326
Mycoplasma genitalium genome sequence
 736, 738
mycosis fungoides 165
myelodysplastic syndrome 160, 163
 chromosome changes 980
 chromosome painting 249
 classification 980
 symptoms 165
myeloproliferative disorders 160
 chromosome changes 980
 classification 981
myotonic dystrophy 10
 trinucleotide repeats 111
N-banding, karyotypic analysis 748, 749
near-isogenic lines (NILs), quantitative
 trait loci mapping 804
neo gene 335–6
neomycin phosphotransferase II 335–6
neonates, cytogenetic analysis 157, 159
nested deletion 514, 523–4, 526
 sequencing projects 524
Netscape software 833, 847
network services 813–35
 bulletin boards 813, 814–15
 e-mail 812, 814
 file transfer 815–33
 Gopher 816, 833
 mail lists 814
 newsgroups 814–15
 Telnet 835
 World Wide Web 817, 833–5
neurofibromatosis type 2 suppressor gene
 442
newsgroups 814–15
nick translation 206–9, 219, 250–2
 DNA labelling with biotin and
 digoxigenin 228–30
NIH/CEPH maps 102
nitro blue tetrazolium (NBT) 604
node 810
non-Hodgkin's lymphoma 162, 165
NotI gene 675–56
nucleolar organizer region (NOR) 747, 748
 staining 155, 156
ocean 432, 434, 435, 436–7
oligonucleotide
 base stacking 568
 fingerprinting 423
 hybridization 524
 labelling with NHS- or ITC-haptens
 612–13
 ligation assay (OLA) 106–7
 multiplex sequencing 524
 ³²P labelling 524
primers
 annealing 569
 design 108–9
 dispersed repeats 110
 purification of labelled 607–8
oligonucleotide-alkaline phosphatase
 conjugate preparation 613–14
oligonucleotide-enzyme conjugate 606–7
 sequence detection 611
oncogenes 656
 denaturing gradient gel electrophoresis
 500
open reading frames (ORFs)
 E. coli 725, 733
 K-12 737
 yeast 700–2, 704, 710
orphan receptor cloning 470
Oryza sativa 776
 see also rice
P1 clones
 DNA extraction 414–16
 end probes 377
 long-range mapping 369, 370–1
 partial digest mapping 377
 screening by hybridization 377–8
P1 cloning
 artificial system 371
 recombinant DNA packaging into phage
 T4 heads 371
P1 library
 Drosophila melanogaster 676, 680, 681–2
 Drosophila viridis 682–3
P1 library construction 401–14
 insert DNA preparation 402–5
 long-range mapping 376–7, 401–14
 P1 packaging 410–14
 recombinant clone recovery 410–14
 recombinant DNA production 407–10
 vector DNA preparation 405–7
P1 maxipreps 377, 415–16

- P1 minipreps 377, 414–15
 P1 vector 377
 system 642
 p53 657
 P element, *Drosophila melanogaster* 673–4,
 679, 680–1
 P element-mediated germline
 transformation in *Drosophila*
 melanogaster 674
³²P-labelled deoxynucleotides 602
³³P-labelled deoxynucleotides 602
pac gene 336–7
 PAC vector system 642
 pachytene karyotyping 748
 palindromic units 728–9
 papilloma transforming proteins E6 and
 E7 660
 papillomavirus 473
PAR1 23
PAR2 23
 parent–offspring pair sampling 36–7
 paroxysmal nocturnal haemoglobinuria
 470, 476
 partial restriction digest mapping
 P1 clones 377
 YACs 374
 paternity analysis 128
 DNA fingerprinting 131, 132
 DNA typing 130, 131
 minisatellite loci typing 133
PAX7-FKHR fusion gene 194
 pCDM8 vector 474, 475
 modifications 474
 transient expression system 472, 473
 pcDNA1 474, 475
 pcDNA3 474, 475
 PCR
 amplification 3
 specific alleles (PASA) 107
 asymmetric 548–9, 570
 colony 287–8
 direct sequencing of products 569–71
 with flow sorting 294
 markers in plant genome analysis 786
 product recovery 552–4
 symmetric 550
 PCR-RFLPs 118–19
 pDUAL 523
 pedigree 9
 breast cancer families 29, 30
 looped 38
 penetrance 10, 28, 31
 probability 36
 Percoll gradient enucleation 354
 perdurance 668
 peripheral blood samples, lymphocyte
 separation 274
Ph1 gene 751–2
 phage
 clones 217
 filamentous 562
 genome size 888
 lambdoid of *E. coli* 728
 lambda-phage clone 521
 phagemids 533, 535, 558
 dideoxy sequencing 563
 DNA preparation 533, 535, 546
 phenocopies 10
 phenotype 9–10
 changes in hybridization 338
 mapping 338–9
 not segregating in Mendelian pattern 23
 phenotypic heterogeneity 33–4
 phenotypic variation 3
 phenylketonuria 655
 Philadelphia chromosome 151, 160
 translocation 246
 phleomycin/bleomycin and phleomycin
 binding protein 337
 photography, automated 168
 photonic noise 309
 photosynthetic proteins, rice 778
 phyletic relatedness 746
 phyllodes tumour of breast 192
 phylogenetic relationship 747
 physical maps 319–20
 integration with genetic 85
 phytohaemagglutinin (PHA) 152
 phytohaemagglutinin-stimulated
 peripheral blood lymphocytes 294
 pixel binning 309, 310
 pJFE14 expression vector 474–5
 plant genome analysis 746–7
 amphidiploids 750–1
 chromosome pairing in hybrids 749–51
 crossability 747
 denaturing gradient gel electrophoresis
 (DGGE) 786
 diploid hybrids 750
 DNA probes 754
 fluorescence *in situ* hybridization 754,
 755
 genomic *in situ* hybridization 754, 755
 in situ hybridization 754
 intergenomic affinity 750
 karyotypic analysis 747–9
 karyotypic features 747–8
 map-based technology 754
 microdissection 755
 molecular markers 754
 molecular tools 753–5
 PCR markers 786
 preferential pairing 750–1
 protein electrophoresis 753
 random amplified polymorphic DNA
 (RAPD) 754–5, 786
 reproductive isolation 747
 techniques 746
 triploid hybrids 750
 plasmid 192, 217
 E. coli 719
 libraries 244
 sequencing vectors 563
 plasmid DNA
 PEG precipitation 540
 sequencing 562
 short alkaline miniprep 532, 533, 540–1
 standard alkaline lysis miniprep 539
 plasmid templates 531–4
 combined anion-exchange
 chromatography/silica gel-based
 purification 533, 534
 PEG precipitation 532, 540
 QIAGEN preparation 532–4
 short alkaline miniprep for plasmid
 DNA 532, 533, 540–1
 silica gel-based purification 533, 534
 standard alkaline lysis miniprep 531–2,
 533, 539
polo gene 668
 poly(A)-RNA preparation 479–80
 polyacrylamide 587
 polyamines 298, 299–300
 polycythemia rubra vera 166
 polyethylene glycol 326–7
 selective precipitation 551
 polymerase chain reaction *see* PCR
 polymorphic markers 3
 polymorphism information content (PIC)
 17, 21, 100
 polymorphisms 3, 47, 99
 A residues 110
 PCR amplified 110
 screening 110
 SSCP analysis protocol 115–17
 substitutional 104, 107
 see also DNA polymorphisms; single-
 stranded conformational
 polymorphism (SSCP) analysis
 polyploid
 heterogenomic 746
 species 746
 population substructuring, allele
 frequencies 130
 positional cloning
 long-range mapping 369
 mouse mutation mapping 640
 Prader–Willi syndrome 242
 pre-banding using Wright's stain 222
 preferential pairing, plant genome analysis
 750–1
 prenatal diagnosis, cytogenetic analysis
 153–4
 primer
 AGK34 227–8
 end-labelled 565, 567
 fluorescent labels 571
 insert-specific 567
 modular 568
 primer walking 514, 520–1, 526, 558
 DNA size 520
 primer synthesis 521
 sequence detection 607, 609
 with short oligomers 568
 priming sites, mobile 521–2
 primitive neuroectodermal tumours of
 CNS 196
 probe competition to remove repetitive
 sequences 416–17
 probe hybridization/sequence-tagged site
 422
 probe-clone incidence matrix 427
 probeorder program 429, 431, 433
 prokaryotic genome sequences 736–8
 prometaphase chromosomes 171–2
 promoter trapping, *Arabidopsis* nuclear
 genome 769
 propionic acidemia 560

- proto-oncogenes 656
 protocols for genetic maps 50–84
 protoplasts 476
 pseudoautosomal region 23, 100
 pseudogenes
 cDNA enrichment 448
 yeast 701–2
 pSP64CS 560
 pSP65CS 560
 pSPL1
 exon trapping 442–3, 445–7, 449–56
 library transient expression 451–2
 recombinant construction 449–51
 pSPL3 exon trapping vector 447
 pSV2 expression vector 334
 puberty 159
 pUC18/19 polylinker 563
 puromycin/puromycin N-acetyl
 transferase 336–7
 PYTHIA program 616–24, 706
- quantitative trait loci mapping 803–4
 quantum efficiency 309
 question and answer dialog 808
 QuickMap 425, 427
 quinacrine banding (Q-banding) 155, 156,
 157, 180–2
- R-banding 152, 155, 158, 182–3
 FISH 224–5
 whole chromosome painting probes
 Plate 4
 5' RACE 464–6
 see also rapid amplification of cDNA
 ends (RACE)
- radiation fusion hybrids
 fingerprint 425
 fragment map 425
 long multiple-copy probe 425
 mapping 424
 single-copy probe target 425
- radiation hybrids 108, 324, 342–4
 analysis 343–4
 applications 323
 mapping 319, 325, 342, 344
 production 358
- radiation mapping 342
 radiation-reduced hybrids 342, 343
- random amplified polymorphic DNA
 (RAPD)
 bulk segregant analysis 785
 plant genome analysis 754–5, 786
 rice genetic linkage map 802
 rice genome mapping 783–5
 sequence-tagged site determination 785
- random amplified polymorphic DNA
 (RAPD)-PCR 113
- random anchor mapping 433–8
 anchored islands 433, 434, 436, 437
 ocean 432, 434, 436–8
 theoretical predictions 435–8
 undetected overlaps 434, 435, 436–8
- random integration 334
- random probing, long-range physical map
 construction 426
- rapid amplification of cDNA ends (RACE)
- 448, 462–6
 ras oncogene 333, 657
 recA mutation 376
 recessive characters, z_2 lod scores 14
 recessive disease, inbred family 38
 recombinant DNA, packaging into phage
 T4 heads 371
 recombinant inbred lines (RILs) 786
 recombinant viral vectors 661
 recombinants 47–8
 recombination fraction 7, 8, 11
 breast cancer 31
 CRI-MAP 62
 linkage analysis 31, 33
 linkage heterogeneity 33
 sex differences 12
 standard error 11
- relational database management system
 (RDBMS) 812
- renal cell adenocarcinoma 193
 renal tumours, media for cytogenetics 196
- Rep protein 719
 REPBAS database 616, 619
- repeat analysis 616
 encoding 626–8
 parsing 626–8
 PYTHIA program 616–24
 recognition of internal repeats 621–4,
 624–5
 recognition of known repeats 616–20,
 624
 repeat subfamily identification 624
- repeats, global 621
- repetitive elements 217
- repetitive extragenic palindrome (REP)
 sequences 728–9
- repetitive sequences, probe competition to
 remove 416–17
- replicons, *E. coli* 719
- representational difference analysis (RDA)
 113
- reproductive isolation, plant taxa 747
- Resource End Database (RED) 53
- restriction enzyme-based fingerprinting,
 Caenorhabditis elegans 688, 689
- restriction fragment length
 polymorphisms (RFLPs) 17, 21, 45
 Arabidopsis 766
 chromosome markers 753, 754
 identification 103–4
 maps of cereal crops Plate 11, 811
 rice 776
 genome analysis 811
 genome mapping 784, 785
 map Plate 11
 typing amplified 118–19
 wheat map Plate 11
- restriction fragment length variant (RFLV)
 635, 636, 638
- retroposons 111
 polyadenylate tract 110
- reverse banding *see* R-banding
- reverse genetics, *E. coli* 725
- reverse transcriptase, enzymatic DNA
 sequencing 565
- Rhodamine 215, 220
- Rhs elements, *E. coli* 728
- ribosomal RNA
 E. coli 726
 genes of *Drosophila melanogaster* 672–3
- rice 776
 cDNA
 analysis 776–8, 779–81, 782–3
 callus proteins 778, 779–82
 clones 776, 778
 root proteins 778, 779–82
 chromosome number 887
 expressed sequence tags (ESTs) 776
 RFLP 776
 map Plate 11
 synteny
 with other cereals 811
 with wheat Plate 11
 rice genetic linkage map 786–804
 construction 801–2
 DNA markers 803, 804
 DNA probes 786–801
 gene tagging for map-based cloning
 802–3
 high-density RFLP 801
 population mapping 786–801
 quantitative trait loci mapping 803–4
 rice genome
 anatomy 810–11
 informatics 809–11
 library screening for genes of other
 cereals 812
 size 776, 804–5, 888
 rice genome analysis
 comprehensive map 809
 data handling 809
 database 809–10
 international federated genome
 databases 810–11
 map-based cloning of target genes 807–9
 RFLP map 811
 YAC library construction 808
 rice genome mapping
 linkage analysis 784–5
 PCR techniques 783–6
 RAPD analysis 783–5
 RFLP markers 784, 785
 single-strand conformation
 polymorphism analysis 785–6
 rice physical map 804–9
 bacterial artificial chromosomes (BACs)
 805
 chromosome isolation 806
 construction strategies 806–7
 cosmid libraries 805–6
 DNA markers 806, 807
 future directions 808–9
 YAC 805–6
 clones 806–7
 rickets, hypophosphataemic 22
 ring chromosomes 157, 159
 RNA isolation 190
 cDNA library construction 478–9
 RNA probes 217
 Rpg1 gene 812
 ³⁵S-labelled nucleotides 602

- sacB* gene 371
- Saccharomyces cerevisiae* 631, 632, 696
- chromosome 696
 - number 887
 - size 888
- clone library 697
- genetic mapping 696–7
- genome 696
- sequencing 513
 - size 888
- model system 696
- see also* yeast
- Sanger sequencing method 561–9
- satellite DNA, *Drosophila melanogaster* 672
- satellited (SAT) chromosomes 748
- Sau*3A cosmid digest 445, 446
- scanning microscopy, sequencing 572
- Schizosaccharomyces pombe* 422
- chromosome number 887
 - genome size 888
 - random anchor mapping 434–8
 - YAC map 428, 430–1
- Secale cereale* banding analysis 749
- secreted proteins, transient expression
- screening 471, 477
- SEG program 624
- segregation, non-independent 6
- selection genes, endogenous 329–31, 332–3
- Sendai virus 326
- seq, genomic 571
- Sequenase 561, 563, 564, 565, 566
- sequence
- characterized amplified regions (SCARs) 785
 - editing 514
 - fidelity 265
 - management programs 514
 - polymorphism conversion to convenient assays 106–7
 - similarity in shared repetitive structure 622
 - with six-phase amino acid translation 514
 - subchromosomal mapping 339
 - tandemly repeated 217
 - transfer 608
- sequence analysis 496, 513
- PCR products 571
 - PCR-amplified segment 104–6
 - programs 514
- sequence detection 600
- charge coupled device 606
 - chemical labelling 609
 - cost 601
 - digoxigenin 611
 - DNA transfer 600–2
 - to membranes 608
- enzymatic labelling 609
- enzyme-linked 605
- chemical end-labelling 607
 - enzymatic labelling 607
 - methods 600, 601, 602–9
 - oligonucleotide labelling 607
 - sequencing protocols 609
- fluorescent 609–10
- handling 601–2
 - hapten-based 611
 - hybridization 600
 - hybridization-based 611–12
 - methods 600–2
 - multiple dye sequence machines 610
 - multiplex sequencing 611–12
 - oligonucleotide–enzyme conjugates 606–7, 611, 613–14
 - primer walking 607
 - project size 601
 - radioactive methods 600, 601
 - silver staining 600, 610–11
- sequence labelling 600
- alkaline phosphatase 603
 - alkaline phosphatase-labelled antibodies 603
 - biotin 602, 603
 - chemiluminescent substrates 604–6
 - colorimetric substrates 603–4
 - cost 601
 - digoxigenin 602, 603
 - 2,4-dinitrophenyl (DNP) 602
 - end-labelled primers 602
 - enzyme-linked methods 600
 - fluorescein 602
 - fluorescent 600
 - fluorogenic substrates 606
 - horseradish peroxidase 603
 - incorporation of labelled nucleotides 602
 - isotopes 602
 - label multiplexing 606
 - oligonucleotides with NHS- or ITC-haptens 612–13
 - radioactive 600, 602
 - streptavidin-phosphatase complex 603
- sequence-tagged site (STS) 422, 424
- assay of *Caenorhabditis elegans* 688
 - content mapping 642
 - Drosophila melanogaster* 678–9, 680–1
 - rice genetic linkage map marker 802
 - rice genome mapping 785
 - YAC contigs 642–3
- sequencing
- automation 571–2, 609
 - by hybridization 568–9, 573
 - Caenorhabditis elegans* 689–90
 - chemical 559–61
 - consistency 561
 - degradation method 558
 - DNA modification analysis 560
 - enzymatic DNA comparison 565–6
 - PCR-amplified products 561
 - vectors 560
- computer resources 514
- cycle 566–7
- dideoxy 514, 558, 560, 562–3
- direct of PCR products 560, 569–71
- DNA
- polymerase 566
 - sequencing machines 626–7
- Drosophila melanogaster* genome 682
- E. coli* 726
- random clone preparation 731
- end-labelled primers 567
- enzymatic DNA 561–9
- 5'-end-labelled primers 565
 - chemical sequencing comparison 565–6
 - Klenow fragment 563–4, 565
 - labelling/termination 563
 - reverse transcriptase 565
 - Sanger 563
 - Sequenase 564
 - Taq* DNA polymerase 564–5
 - techniques 563–6
- enzymatic method 558
- gel electrophoresis 572
- high resolution denaturing
- polyacrylamide electrophoresis 558
- in vivo* amplification methods 531–6
- large-scale of *Drosophila melanogaster* 681
- mass spectrometry 572–3
- membrane development 608–9
- multiplex 524–6, 559
- nested deletions 523–4
- novel techniques 566–9
- PCR technology 559
- primer walking with short oligomers 568
- primer-directed 567–8
- rice 776–8
- robots 777
- scanning microscopy 572
- shotgun 518–20, 526
- solid phase 559, 570
- strategies 514, 518
- directed 520
 - nested deletion 526
 - ordered 518
 - primer walking 520–1, 526
 - random 518
 - shotgun 518–20, 526
 - transposon mediated 526
- technology 514
- template
- amplification/purification 531
 - PCR products 558
 - preparation 514
- template generation 569–71
- asymmetric PCR 570
 - automation 572
 - cycle sequencing of PCR 571
 - double-stranded 570–1
 - lambda exonuclease-generated single-stranded DNA 570
 - solid-phase 570
- transposon-facilitated 681
- transposon-mediated 521–2, 523
- walking primers 520–1, 526, 558
- sequential digestion 523
- severe combined immune deficiency (SCID) gene therapy 653–5
- sex chromosome
- cytogenetic analysis of abnormality 159
 - locus inheritance pattern 100
- sex linkage 22–3
- sex reversal 345
- Sézary's syndrome 165
- shaker-1* gene 641
- short interspersed repeat elements 344

- short tandem repeats 21
 see also microsatellites
- shotgun sequencing 514, 518–20, 526
 assembly 519
 contigs 519
 directed 519–20
 editing 519
 library preparation 518–19
 sequence acquisition 519
- silicon intensified camera 308
- silver staining, sequence detection 610–11
- simple regions 621, 623
- simple sequence length polymorphism (SSLP), segregation analysis in mouse 635, 636
- simple tandem repeats
 DNA typing 129
 PCR typing 135–6
- single-copy probe hybridization, data 427–30
- single-stranded conformational polymorphism (SSCP) analysis 103, 104, 105
- denaturing gradient gel electrophoresis 503
- polymorphism 115–17
- rice genome mapping 785–6
- sensitivity 503
- single-stranded DNA binding protein 521, 568
- slab gel electrophoresis 579–85
 apparatus 582
 automated sequencing 585
 autoradiography 584
 band width 582
 bis-acrylamide 580, 581
 blotting 584–5
 buffer 581
 capillary blotting 584
 catalyst systems 581
 crosslinkers 580
 degassing 581
 denaturants 581–2
 direct transfer electrophoresis 584–5
 electric field conditions 582–3
 electroblotting 584
 gel composition 580
 gel dimensions 582
 gel matrix 579–81
 gradient gels 583–4
 Joule heat 582
 loading 582
 manual sequencing 579–85
 polyacrylamide chains 580, 581
 sample wells 582
 sharktooth comb 582
- SMPL program 621–4, 626
 algorithms 626–8
- solid tumour cytogenetics 189–90
 CGH 194–5, 196, 206–9
 technique 189
 chromosome
 direct preparation 198–9
 harvesting 193
 preparation 190–3
 rearrangements 189
- culture media 190
- direct preparations 191, 198–9
- disaggregation 190–1, 196–8
- fine-needle aspirates 190
- FISH technique 189, 193–5
- harvesting
 by removal of adherent cells 204–5
 in situ 205–6
- Lipofectin-mediated transfection 202–3
- long-term cultures 191–3, 202–3
- media 196
- nick translation labelling of genomic DNA 206–9
- passaging cells 203–4
- RNA isolation 190
- short-term cultures 191
 from cell suspensions 199–201
 from explants 201–2
- tumour imprint preparation 209
 pretreatment prior to hybridization 209–10
- tumour sample 190
- washing 190–1
- solid tumours
 chromosome rearrangements 982–3
 gene amplifications 984
 karyotypes 189
 unsynchronized cultures 277
- solutions 863–7
- somatic cell hybrids 107–8, 323
 auxotrophic mutants 333–4
 biology 324
 cell culture 325–6
 cell source 327–8
 characterization 344–5
 chromosome segregation 328
 cloned DNA mapping 325
 dominant selectable marker insertion into mammalian genome 334–8
 donor cells 328
 donor chromosomes 357
 endogenous selection genes 332–3
 enucleation
 from plastic bullets 351–3
 Percoll gradient 354
 genotype characterization 325
 half-selection 333
 HAT preparation 346–7
 HAT selection 332–3
 interspecific 324, 325
 irradiation and fusion gene transfer 342–4
 long-range mapping 369
 marker analysis 344–5
 microcells
 filtration 354–5
 fusion to whole recipient cells 355–7
 micronucleation optimization 349–50
 MMCT-generated monochromosomal 334, 335
 morphology 327–8
 parental cell selection 328
 phenotype mapping 324–5
 positional cloning of human disease genes 325
- radiation hybrid production 358
- recipient cell lines 328
- selection 328, 329–31, 332–4
- types 324
- whole-cell fusion 338–9, 347–9
- SOX9 gene 345
- SP6 promoter element 560
- spheroplasts 476
- spinal muscular atrophy, trinucleotide repeats 111
- Standard Query Language (SQL) 812
- startle disease 641
- strain distribution pattern, mouse genome mapping 637
- streptavidin bridge 525
- streptavidin-phosphatase complex, sequence labelling 603
- subarray sampling 309
- subchromosomal mapping of DNA sequences 339
- subcloning, nest deletion 568
- subgenomic libraries 107–8
- substitutional polymorphism 104, 107
 analysis 108
- supF* suppressor tRNA gene 522
- surface proteins, transient expression screening 471
- susceptibility allele 37
- SV40 473
 early region 192
 promoter 334
 vector for cDNA library construction 473, 474
- SV40-based plasmids 475
- SWISS-PROT database 736, 771
- synchronization technique for bone marrow samples 184–5
- Synchronous Multimegabit Data Services (SMDS) 811
- synkaryons 324
- synovial sarcoma, media for cytogenetics 196
- synteny 811–12
 map Plate 11
- System for Integrated Genome Map Assembly (SIGMA) 83
- T7 polymerase 521
 sequence analysis of PCR products 571
- t(9;22) 160
 translocation 151, 167
- T-cell receptor β heterodimer 471
- T-DNA, *Arabidopsis* nuclear genome 767–8, 769
- Ta1 element 763
- tandem repeat
 sequences 217
 variability 108–9
- Taq DNA polymerase 561
 sequence analysis of PCR products 571
 sequencing 564–5, 566
- Taq polymerase 106, 521, 558, 565
- Target End Database (TED) 53
- target genes, map-based cloning 807–9
- Tat1 element 763
- tat intron 446

TAT protein 660
Telnet 835
telomere
 banding 155
 trap cloning 699
 yeast 699, 705–6
temperature, melting (T_m) 496, 498, 499
temperature gradient gel electrophoresis,
 DNA sequence analysis 496
template amplification 531
 agarose gel electrophoresis 533, 536–7,
 538
 asymmetric PCR 533, 536, 537, 548, 549
 biotin-streptavidin system 537, 551–2
 column purification 533, 538
 cosmids 533, 535–6, 546–7
 detergent extraction for M13 DNA
 542–3
 DNA direct sequencing in LMP agarose
 533, 538, 555
 double-stranded DNA sequencing
 template purification 537, 550
 freeze and squeeze method 533, 538,
 552–3
 in vitro methods 533, 536–8
 M13 templates 533, 534–5
 magnetic bead purification of M13 DNA
 543–4
 PCR product direct sequencing 536
 PCR product recovery
 agarose method 554–5
 column purification 553–4
 freeze and squeeze method 552–3
 PCR-amplified material molecular
 cloning 536
 PEG precipitation
 plasmid DNA 540
 selective 533, 537, 551
 phagemids 533, 535, 546
 phenol/chloroform extraction 533, 538
 plasmid templates 531–4
 pure double-stranded 537–8
 short alkaline miniprep for DNA 540–1
 silica gel-based purification 533, 538,
 545
 single-stranded DNA sequencing
 template generation 537–8, 548,
 549, 551–2
 standard alkaline lysis miniprep of
 plasmid DNA 539
 standard PEG-phenol method for DNA
 recovery from M13 phage 541
 symmetric PCR 533, 536, 537–8, 550
template DNA of rice 777
template purification 531
N,N,N',N'-tetramethyl-1,2-diaminoethane
 (TEMED) 581
tetranucleotide repeats 111–12
Texas red 215
 biotin detection 235–6
TFASTA program 771
 α -thalassaemia 242
 β -thalassaemia 500
thalassaemia, gene therapy 654, 655
thermal cycle sequencing 566–7
Thinopyrum bessarabicum, chromosome
 pairing 751, 752
thymidine
 block synchronization 171–2
 cell synchronization 265–6
thymidine kinase 332–3
tissue plasminogen activator
 repeats 619
 sequence 616
tissue-specific enhancers 655
tk gene 338
Tn3 transposon family 521
Tn5 transposon family 521
toxicological assay of chemicals 295
transduction, generalized 724
transformation, YAC cloning 372
transient expression screening 470–2
 advantages 471
 cDNA library construction 472–6
 cDNA synthesis 472
 disadvantages 471–2
 extracellular proteins 477
 intracellular proteins 477
 methods 476–7
 procedure 470–1
 screening ligand 472
 secreted proteins 477
 supernatant bioassay 477
 for surface molecules by panning and
 rescue 476, 486–91
 vectors 472–6
translocation breakpoints, microdissection
 273
translocations
 balanced 156
 unbalanced 156, 157
Transmission Control Protocol/Internet
 Protocol 810
transmission distortion tests 40
transposable elements of *Drosophila*
 melanogaster 673
transposon tagging
 Arabidopsis nuclear genome 768–9
 Drosophila melanogaster 674
transposon transmission in *E. coli* 728
transposon-mediated sequencing 514,
 521–2, 523, 526
 mobile priming sites 521–2
 transposon-generated deletions 522,
 523
TRAP gene mapping 221
trinucleotide repeats 47, 111–12
triploid hybrids, plant genome analysis
 750
trisomies 156–7
Triticum aestivum, chromosome pairing
 751–2
Triticum turgidum, C-banding 749
trpB gene 337
tryptophan/tryptophan synthase
 (β -subunit) 337
Tth1111 site 560
tumour antigen expression 658
tumour cell transplantation into nude mice
 193
tumour necrosis factor 659
tumour suppressor genes 656
 denaturing gradient gel electrophoresis
 500
tumour-infiltrating lymphocytes (TILs)
 659
tumourigenesis 159
twins, zygosity testing 131
t(X;18)(p.11.2;q11.2) 194
Ty elements, yeast 702, 706
tyrosine kinase 641
undetected overlaps 434, 435, 436–8
universal DNA amplification 269
universal relative locator (URL) 833
UNIX 50–1, 838–41
 activities 838–40
 background/foreground processing
 839–40
 case 838
 commands 838
 comparison with VMS 840
 control-key combinations 841
 deleting 838
 file names 839
 output control 840
 program execution control 839
 Wildchar characters 840
 working in other directories 839
upstream acting sequences (UAS) 703
upstream repressing sequences (URS) 703
variable number of tandem repeats
 (VNTR) 108–9
vectors
 adenoviral 662
 delivery 661
 development 662–3
 DNA preparation for cosmid libraries
 375
viral genes 192
viral infection 659–61
viral vectors 662–3
virus, genome size 888
viscotoxin 778
VMS 840–1
Waldenström's macroglobulinaemia 166
Webcrawler 820, 834
wheat
 chromosome markers 802
 chromosome number 887
 RFLP map Plate 11
 synteny with rice Plate 11
white gene 668, 669
whole-cell fusion 338–9
 mammalian cells 347–8
whole-cell hybrids 324, 338–9
 applications 323
whole-chromosome libraries 248
whole-chromosome map construction 342
whole-chromosome paints 246, 247
wide area networks 810, 811
Wilm's tumour suppressor 780
World Wide Web 805, 817, 833–5
 databases 847–61
 genome resource centres 847–61
 URLs for live help 841

- Wright's stain 222
- X chromosome 22
centromeric probe Plate 8
genetic recombination 23
X-specific part 100
- X-linked agammaglobulinemia (XLA) 641
- X-linked traits 22
- X-PED 53
- X-Phos *see* 5-bromo-4-chloro-3-indolyl phosphate (BCIP)
- X-Y pairing region 100
- Xa-1* gene 807
- xanthine phosphoribosyltransferase 336
- xeroderma pigmentosa 470, 475, 476
- Xho*I restriction site 270
- Y chromosome, inverted 159
- YAC-to-YAC hybridization 425, 426
- YACs 108, 147
agarose block preparation 373, 391–3
anchored framework map construction 643, 644
cDNA enrichment 443
chimaerism 432
chromosome walking 374
clones
end-specific probe generation 395–7
replication 373
rice physical map 806–7
cloning 372
combination with chromosome paints 243
contig 108, 422
assembly 273
construction 271
direct cDNA isolation 447
direct hybridization to cDNA filters 456–7
DNA partial restriction digest mapping 394–5
expressed sequence isolation 455
exon-trapping 447
filters 429
FISH 218
Génethon human genome map 425
inner product mapping 244–5
insert size 369
library
Arabidopsis 765–6
arraying 372, 389
chimaeric clones 642
construction 372, 381–8
Drosophila melanogaster 675–56
filter lifts 373, 389–91
mouse 642, 643
rice 808
screening 373
long-range mapping 369, 370, 381–8
map for *Schizosaccharomyces pombe* 428, 430–1
P1 clones 371
partial restriction digest mapping 374
PCR 374–5
physical genome maps 641–2
pooling 373
preliminary characterization 373–4
preparation
by ligation 384–5
for transformation 386
probe
generation 374
ordering 429
rice physical map 805
screening by hybridization 377–8
size
determination 373
fractionation by pulsed-field gel electrophoresis 393–4
subcloning 370
use 372–5
vector 372
arm preparation 382–3
yeast spheroplast preparation 386–7
- yeast
Arabidopsis genes cloned 767
ARS elements 705, 709
base composition 704–5
chromosome
breakage 709
sequence homology 706
telomeres 699
chromosome II repeat sequences 706, 707
clusters
duplicated genes 708
homology regions 708–9
duplicated genes 708
functional analysis 711
GC content 704–5
- gene
density 704–5
organization 702–7
genetic map 707
genetic redundancy 707–9
genome 696
architecture 702–7
organization 707–11
genome project 696, 697–700
chromosome sequencing 697–9
cloning 697–9
mapping 697–9
nested chromosomal fragmentation 699
quality control 699–700
sequence analysis 74
sequence assembly 699–700
sequencing strategies 699–700
strategy 697
human connection 709–10
information sources 714
intergene intervals 704
introns 701–2
open reading frames (ORFs) 700–2, 704, 710
physical map 707
proteome 700–1
pseudogenes 701–2
putative membrane proteins 702
putative mitochondrial proteins 702
repeat sequences 706–7
sequence variation among strains 709
spheroplasts
preparation 386–7
transformation 388
telomeres 705–6
transcriptional unit arrangement 703
transformation efficiency 697
Ty elements 702, 706
upstream acting sequences (UAS) 703
upstream repressing sequences (URS) 703
see also Saccharomyces cerevisiae
yeast artificial chromosomes *see* YACs
- z_1 lod scores 13–14, 17
tables 19
 z_2 lod scores 14–16, 17
tables 19–21
zygosity testing, twins 131

ICRF Handbook of Genome Analysis

The *ICRF Handbook of Genome Analysis* is a combination of protocol manual and informational resource, with expert contributors drawn from a wide range of research centres. It describes and evaluates a wide range of techniques, providing step-by-step protocols. The two volumes cover both the human genome and genomes of other model organisms.

Contents

Volume 1

Section 1: Genetic mapping

Section 2: Physical mapping 1: cytogenetic analysis

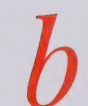
Section 3: Physical mapping 2: long-range mapping and gene isolation

Volume 2

Section 4: DNA sequencing

Section 5: Other model systems

Section 6: Internet resources



Blackwell
Science

ISBN 0-632-03728-8



9 780632 037285 >